2014

Estimation Under Stochastic Differential Equations

Shan Yang
Iowa State University

Estimation under stochastic differential equations

by

Shan Yang

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Song Xi Chen, Major Professor

Karin Dorman

Arka Ghosh

Ananda Weerasinghe

Huaiqing Wu

Cindy Yu

Iowa State University

Ames, Iowa

2014

DEDICATION

To My Family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to my major professor Dr. Song Xi Chen for his patient guidance, inspiring suggestion and constant encouragement throughout this research and the writing of this thesis. I am very thankful to my committee members, Dr. Karin Dorman, Dr. Arka Ghosh, Dr. Ananda Weerasinghe, Dr. Huaiqing Wu, and Dr. Cindy Yu for their precious help. Thanks also go to my parents and all my friends for their support and inspiration. Special thanks to Jinyuan Chang for help coding and running Mathematica to justify the results in the example of two dimensional O-U process.

ABSTRACT

Stochastic approaches are used in modern financial analysis to explore the underlying dynamics of securities like stocks and options. Statistical modeling and inferences within this aspect is an important concern because pricing errors could lead to serious economic losses. In this thesis, statistical estimation motivated by real applications are developed for inferences under stochastic diffusion processes using tensor method and kernel smooth method.

We consider in Chapter 2 parameter estimation for multi-factor stochastic processes defined by stochastic differential equations. The class of processes considered are multivariate diffusion which are popular processes in modeling the dynamics of financial assets. We quantify the bias and variance by developing theoretical expansions for a large class of estimators which includes as special cases estimators based on the maximum likelihood, approximate likelihood and discretizations. We apply the proposed methods to evaluate bias in estimated contingent claims. We also provide simulation results for a set of popular multi-factor processes to confirm our theory.

Our Chapter 3 is dedicated to improve the estimation of the market volatility, specifically the VIX index introduced by Chicago Board Option Exchange (CBOE). This index provides a way to measure the 30-day expected volatility of the S & P 500 index. Among a few ways to estimate it, the CBOE and the Goldman Sachs had developed an estimator based on the concept of fair value of future variance. In realizing the discretization error, truncation error, and the approximation error in their estimator, as well as the possible option pricing errors involved, we propose a new method that combines the CBOE method and the kernel smoothing method. We derive the weak convergence property of our estimator. Simulation is run to justify the improvement.

CHAPTER 1. GENERAL INTRODUCTION

Long long ago, our ancestors have their goods or services exchanged between each other directly to satisfy their living needs. With the development of some financial mediums like money, people are facilitated to trade more diversified goods with various parties. During these activities, people are facing some uncertainties. For example, they may trade the same amount of their products at different times or with different parties for different amount of money. We say the traders are facing risks.

How to maximize one's profit under certain amount of risk is one of the main targets of modern financial analysis. Various assets, tools, and models are developed to help achieve this purpose. In this chapter, we briefly introduce some related concepts that motivate our research. More rigorous mathematical definitions will be given in later chapters.

1.1 Financial Market

Risk is one of the key concepts in modern financial world. A **riskless asset** is an asset with deterministic future value. For example, if you lend out $100 today and will be paid $105 one year later with 100% probability, you are holding a riskless asset. However, if the future value of the asset is subject to change due to various factors, people, environment, etc., it is a **risky asset**. In our example, if your future payment after one year is not assured to be a fixed amount, but may vary depending on the borrower's financial circumstance, you are holding a risky asset.

A **stock** is a share of a company that entitles the the holder a fraction ownership of that company. These companies are usually public limited companies, which means the owners of such a company have no liability for the company's debts if it bankrupts. The stock owner earns money when the company pays out dividend or the stock price increases.

An **index** is the weighted average of the prices of selected representative stocks. The corresponding weights and stocks are subject to change. An index can be used to represent the overall performance of the stock market. For example, the **S & P 500** index, or the Standard & Poor's 500, is a stock market index based on the market capitalizations of 500 diversified large companies' common stocks. The **S & P 100** index contains 100 leading U.S. stocks which are a subset of these that constitute the S & P 500.

A **future** is a standardized contract signed by two parties today to buy or sell a specified asset at a fixed future date with the amount of the asset and the delivery price agreed today. It is a legally binding that links both parties to an obligation of a future delivery. A future contract itself is with no current value.

An **option** is a contract that gives its buyer or owner the right, but not the obligation, to buy or to sell a certain amount of financial asset, like stock or index, on or before a certain date at a specific price. The date is called "maturity" and the price is called "strike price" or "exercise price". If the contract can only be exercised at maturity, it is called a **European option**. An option to buy is a **call option**, while an option to sell is called a **put option**. The seller of the option will ask for a certain amount of money, called **option price** or **premium**, from the buyer for obtaining the corresponding right. When one expects the stock price to rise, a call option will be bought or a put option will be sold. When the strike price is equal to the current underlying asset price, the option will be called as **at-the-money**. An option is **near-the-money** if the strike price is close to the current underlying asset price. When a call option is with strike price higher than the market price of the underlying asset, or when a put option's strike price is lower than it, we call the corresponding option as **out-of-the-money**.

A **derivative** is a financial contract which derives its value from the performance of the underlying asset, such as index or interest rate. It includes a variety of financial contracts such as futures and options. A **security** is a tradable asset of any kind, such as stocks, futures and options. A **contingent claim** is a claim that can be made when certain outcomes occur.

The **security return** is the gain or loss of a security in a particular period. Specially, an **interest rate** is the rate at which the borrower pays the lender for the use of the money at the end of a pre-specified period. In the riskless asset example we just described, the **annual interest rate** is $r = \frac{105-100}{100} * 100\% = 5\%$. If we assume the interest rate is distributed and accumulated instantaneous, the **instantaneous interest rate** will be $R = \ln(1+r) = \ln(1+5\%) \approx 4.88\%$. For small r and R, the difference between them will be small.

Suppose an asset, say a stock, is with price S_t, where $0 \leq t \leq T$. When the instantaneous interest rate is a constant, the **discounted price** of this asset at time t is defined as $e^{-R(T-t)}S_T$.

Transaction fee is a fee that one pays to buy or sell assets. It is different from the asset price. **Trading volume** is the total quantity of certain assets bought and sold during a given period. **Net buying pressure** is the difference between the number of buyer-motivated contracts traded each day and the number of seller-motivated contracts traded each day.

Volatility is the square root of the variance of the price change. The larger the volatility, the higher the possible profit. There are various volatility estimators. **Historical volatility** refers to the standard error of the log asset ratios. **Implied volatility** is the model implied volatility that corresponds to a given option price. **Integrated volatility** is essentially the quadratic variation of the log asset price process. The Chicago Board Options Exchange, together with Goldman Sachs, developed a model free method to estimate the expected volatility index, VIX. Volatility index **VIX** estimates the future 30-day volatility of the S & P 500 index. The major part of final adapted estimator is the weighted sum of the S & P 500 put prices and call prices over a wide range of strike prices.

In summary, with the original debut of the stock market that can be traced to as early as 12th century, companies started to use stocks to raise money. To satisfy the need of different investors, various derivatives like options are created. Rational investors are eager to profit by buying stocks, options, and so on at cheaper prices while selling them at higher prices, while being exposed to as low risk as they can. In order to see the trend of the stock market in general, stock market indexes such as S & P 500 are introduced. Volatility index is also introduced to help estimate the market volatility.

1.2 Stochastic Differential Equations

In order to model the dynamics of financial derivative, lot of models are introduced. Most of those popularly used are stochastic differential equations driven by Brownian motions. The general form of a stochastic differential equation and some interesting special cases are introduced in this section. Important model specific properties will be illustrated in later chapters. Please refer to Stroock and Varadhan (1979), Karatzas and Shreve (1987), or Øksendal (1998) for more details.

Brownian Motion Botanist Robert Brown first discovered that the pollen suspended in water is moving along a random trajectory with the aid of a microscope in 1828. Seventy seven years later, Albert Einstein explained that the dispersal or diffusion of the pollen in the water is resulted by random collisions of the pollen with individual water molecules. An adapted stochastic process, $B = \{B_t, \mathcal{F}, 0 \leq t < \infty\}$ defined on some probability space $(\Omega, \mathcal{F}, P)$, called Brownian motion (BM) is developed to describe such random behaviors:

1. $B_0 = 0$ a.s.,

2. $B_t - B_s$ is independent of $\mathcal{F}$ and $(B_t - B_s) \sim N(0, t-s)$ for $0 \leq s \leq t$).

where $N(0, t-s)$ denotes the normal distribution with expected value 0 and variance $(t-s)$. We can see BM is almost surely continuous and $E[B_t^2] = t$. Brownian motion is also called as Wiener process in honor of Norbert Wiener, who proved the non-differentiability of the paths.

The Brownian motion is one of the simplest continuous time stochastic processes. It's widely adopted to describe the fluctuation of stock price, the thermal noise in electrical circuits, etc.

A multi-dimensional Brownian motion is defined similarly. To be concise:

1. $P(B_0 \in \Gamma) = \mu(\Gamma)$, $\forall \Gamma \in \mathcal{B}(\mathbb{R}^d)$;

2. B_t is a Gaussian process, i.e. for all $0 \leq t_1 \leq \ldots \leq t_k$ the random variable $Z = (B_{t_1}, \ldots B_{t_k}) \in \mathbb{R}^{nk}$ has a (multi)normal distribution;

3. B_t has independent increments, i.e. $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$ are independent for all $0 \leq t_1 < t_2 < \ldots < t_k$.

Stochastic Differential Equation The stochastic differential equation (SDE) was first suggested by Paul Lévy as an alternative representation for diffusion process, which essentially means a Markov process with continuous sample paths and can be characterized by an operator, infinitesimal generator. The theories about stochastic differential equation have been well developed by Kiyoshi Itô, etc.

A general form of a continuous, adapted d-dimensional process $\{X_t\}$ which exists under certain regular conditions on a probability space $(\Omega, \mathcal{F}, P)$ can be given as:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t, \tag{2.1}$$

where $\{B_t, \mathcal{F}_t\}$ denotes a d-dimensional standard Brownian motion, the coefficients, the drift function μ and the diffusion function σ, are Boreal measurable. This equation should be interpreted as an informal way of expressing the corresponding integral equation

$$X_{t+s} - X_t = \int_t^{t+s} \mu(X_u, u)du + \int_t^{t+s} \sigma(X_u, u)\, dB_u.$$

Please note that the integration with respect to Brownian motion part is under Itô integral sense, which is different from the Riemann integral.

Black-Scholes Model Black and Scholes (1973) suggested to model the underlying stock price by the following process

$$dS_t/S_t = rdt + \sigma dB_t,$$

where $r, \sigma \in \mathbb{R}^+$, r is the risk free interest, and $\{B_t\}$ is the standard Brownian motion. The law of S_t is lognormal. This simple model mimics the realistic stock price process relatively adequately, and hence is well adopted in the financial world.

Merton Jump Model In order to better describe the financial market when there is sudden price change, Merton (1976) first introduced the so called Merton jump model. Assume the dynamic of stock price can be described by

$$dS_t/S_t^- = (r - \lambda(m-1))dt + \sigma dW_t + (J_t - 1)dN_t,$$

where $r, \lambda, m \in \mathbb{R}^+$, r is the risk free interest rate, σ^2 is the instantaneous variance of the return, $\{W_t\}$ is the standard Brownian motion, $\{N_t\}$ is a Poisson process with parameter λ, and $\{\log(J_t)\} \overset{i.i.d.}{\sim} N(\log(m+1) - \frac{1}{2}v^2, v^2)$. The Brownian motion $\{W_t\}$, jump size $\{J_t\}$ and jump frequency $\{N_t\}$ are all independent of each other.

Mean Reversion Process A special class of diffusion process called mean reversion process has caught special attention in the financial world. For example,

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dB_t,$$

where $\{B_t\}$ is a standard one dimensional Brownian motion, the mean reversion rate κ, long-term mean or mean reversion level α, the volatility σ, and the ρ belongs to $\mathbb{R}^+$. It nests the famous Vasicek model or called O-U process (when ρ=0) and CIR model (when ρ=0.5). The asset price, like constantly changing stocks' prices, are believed by some to tend to move to the average price over time. Bessembinder, Coughenour, Seguin and Smoller (1995) used the term structures of futures prices from agricultural comodities, crude oil, metals, and financial asset prices to empirically proved the existance of mean reversion in each market. Bessembinder, Coughenour, Seguin and Smoller (1995) used the term structures of futures prices from agricultural comodities, crude oil, metals, and financial asset prices to empirically proved the existance of mean reversion in each market.

1.3 Parametric and Nonparametric Estimation

Estimating Equations Estimating equations is used as an alternative way to estimate model parameters. It can be thought as a generalization of the method of moments, least squares, and maximum likelihood. The essential idea is, when given a model, adopt some known functions $g(x_{d\times 1}, \theta_{p\times 1}) : \mathbb{R}^{d+p} \to \mathbb{R}^r$ to restrict the relationship between the sample data X and the parameter θ. Then, attain estimation to the parameter θ through solving the equation, say $g(X,\theta) = 0$. Please refer to Hardin and Hilbe (2003) for more details.

Tensor Notation What is a tensor and what's the use of the tensor method? There's no succinct answer. Let us assume an example. For a linear function

$$a = a_1 v^1 + a_2 v^2 + \ldots + a_n v^n,$$

it can be written abbreviated as $a = a_i v^i$. Motivated by modern applications in higher dimensions, we strive to work under these succinct notations to achieve beautiful results.

Infinitesimal Generator Denote $C^2(\mathbb{R}^d)$ to be the set of real-valued twice continuously differentiable functions. The unique solution to the stochastic differential equation (2.1), if exists, is a realization of the stochastic processes which satisfies

$$A_t f(x) = \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 f}{\partial x^2} + \mu(x,t)\frac{\partial f}{\partial x}, \ \ \forall f \in C^2(\mathbb{R}^d),$$

where the differential operator A is called the infinitesimal generator of the diffusion process $\{X_t\}$. For the relation between the process and the corresponding generator and more related properties, please refer to Karatzas and Shreve (1988). Because under certain regularity conditions, we will prove the conditional expectation of a function of X_t can be approximated by the sum of a series of generators applied to it, we can use the infinitesimal generator to help find the conditional expectation when explicit form of the transitional density is unknown. We illustrate how to apply this method under the one dimensional O-U process in equation (5.14) in section 2.5.2.

Parametric Estimation Under Stochastic Differential Models Extensive papers have been dedicated to explore pricing and estimation problems under stochastic differential models. Under the mean reversion model, documented simulation results shows that the drift parameters are estimated less accurate than others. Tang and Chen (2009) derived the high order bias for one dimensional maximal likelihood estimators of Vasicek and CIR model. Wang, Phillips and Yu (2010) gives the bias under the exact discreet time model suggested by Philips (1972). Based on those papers, we get better understanding of how higher order bias influences the whole estimation result. We explore a new method that can be applied to multivariate stochastic differential models for deriving estimation biases and variances based on general estimating equations. Some examples of applying this method and some simulation results will be shown too.

Kernel Method In non-parametric statistics, a kernel is a continuous, bounded and symmetric real function that integrates to one:

$$\int_{-\infty}^{\infty} K(x)dx = 1.$$

It is essentially a weighting function used in non-parametric estimation techniques. Popularly used kernel functions include uniform, triangle, Epanechnikov, biweight, tricube, and Gaussian. Please refer to Härdle (1989) for more details.

N-W Estimator When no model is available, or no model is believed to be true, or when the size of data is very large, we tend to let the data to speak for themselves. Motivated by the conditional expectation form, Nadaraya (1963) and Watson (1964) proposed the N-W estimator for the regression. N-W Estimator is a modern nonparametric method of statistical inference, which become more popular with the boom of computer. It is essentially the weighted sum of sample values with most weights usually concentrate around the nearby points:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K(\frac{x-X_i}{h})Y_i}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})},$$

where h is called the bandwidth and the K is a kernel function.

Bandwidth Selection The bandwidth h of the kernel is a free parameter whose chosen value can strongly influence the estimation result. Hence, careful selection of the bandwidth is highly advised. It is often recommended to choose an optimal bandwidth to minimize some global error criterion automatically so as to get a good starting point. For example, the bandwidth can be selected to minimize the mean square error of the N-W estimator, or to be the one that minimizes the cross-validation target or the penalizing function. Please refer to Härdle (1989) for more discussions.

Kernel Estimation for VIX It is of special interest to estimate the market volatility based on timely updated market prices. Based on the work in Demeter, Derman, Kamal and Zou (1999) about variance swap, CBOE white paper (2009) specifies a very practical way to calculate the estimated VIX value. Some other estimators are proposed by Britten-Johns and Neuberger (2000), Jiang and Tian (2007), etc. To reduce the VIX estimation bias, we explore the possible sources and we believe option pricing errors add more bias to the VIX pricing. We suggest adding a kernel smooth step to the VIX estimation procedure. Both theoretical and simulation results support our assumption that the estimation bias will be reduced under our targeted cases when the kernel smooth is introduced.

1.4 Dissertation Organization

This dissertation consists of two parts. The first part, Chapter 2, is about parameter estimation for multi-dimensional stochastic differential equations driven by Brownian motion based on tensor method. In the second part, Chapter 3, we try to modify the standard method adopted by the Chicago Board Options Exchange (CBOE) to estimate the volatility index of S & P 500, VIX, by adding a non-parametric estimation step.

CHAPTER 2. PARAMETER ESTIMATION FOR MULTIVARIATE STOCHASTIC DIFFERENTIAL EQUATIONS

Given a set of $d-$dimensional financial data sampled with a time interval $\triangle$,

$$X_0, X_\triangle, ..., X_{n\triangle}$$

which are assumed to be observed values of a $d-$dimensional stationary stochastic process $\{X_t\}$

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dB_t, \tag{0.1}$$

with infinitesimal generator $A(\theta)$:

$$A(\theta)f(x) = \sum_{i=1}^{d} b_i(x,\theta)\frac{\partial}{\partial x_i}f(x) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}(\sum_{k=1}^{d}\sigma_{ik}(x,\theta)\sigma_{kj}(x,\theta))\frac{\partial^2}{\partial x_i \partial x_j}f(x),$$

where $b(x,\theta) = (b_1(x,\theta),\ldots,b_d(x,\theta))^T$, $\sigma(x,\theta) = (\sigma_{ij}(x,\theta))_{d\times d}$, and $\theta_{p\times 1}$ is a vector of length p. We are interested in knowing how accurate we can estimate the model parameters θ based on observed historical data, which may be used to derive the price of options, bond prices, and other financial derivatives. Without special explanation, from now on we use X_i to denote $X_{i\triangle}$ when samples are mentioned, to simplify notation.

To make our theory generally applicable, we develop a theory without restricting the estimator to be any specific one, like the maximal likelihood estimator or the martingale estimating equation estimator. We adopt the estimation equation approach. Suppose we have an estimating equation g which satisfies [1]

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i, X_{i+1}; \theta) = 0, \text{ and } E[g(X_i, X_{i+1}; \theta_0)|X_i] = 0.$$

Denote the implied estimator for θ_m as $\hat{\theta}_m$, $1 \le m \le p$. Then we obtain the major orders of the this estimator through tensor expansion. We prove the remainder terms are ignorable under some fair assumptions.

After we attain the general theoretical forms of the bias and variance for the interested parameter θ, we deduce the specific form of bias and variance for some interesting processes. A stochastic diffusion process with realization

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t dB_t$$

is called an Ornstein-Uhlenbeck (O-U) process. Here, $\{X_t, t \ge 0\}$ is a $n-$dimensional stochastic process driven by a $n-$dimensional Brownian motion $\{X_t, t \ge 0\}$, $\kappa_{n\times n}$ is a constant matrix, $\alpha_{n\times 1}$ and $\sigma_{n\times 1}$ are two constant vectors. Similary, a CIR process is defined to satisfy

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t.$$

More details about O-U processes and CIR processes will be provided in the Section 2.1.1. We derive the result under both O-U and CIR processes and conduct extensive simulation study to confirm our results and try a parametric bootstrap method to reduce the bias. Write $\theta = (\alpha, \kappa, \sigma^2)$. The vector α represents mean reversion level, the vector κ denotes the mean reversion rate, and the matrix σ^2 stands for volatility. We notice the order of bias and variance for two dimensional O-U process and CIR process are of the same order as their one dimensional counterparts: $O(1/T)$ for mean reversion rate κ, $O(1/n)$ for volatility σ, and $o(1/n)$ for mean reversion level α. Here $T = n\triangle$ denotes the time of observation. To be intuitive, the relative bias for the mean reversion rate κ can be as large as 200% with 100 sample, or reduced to as small as a few percent when samples increase to near 5000.

[1] Let ϖ be a constant. If $E[g(X_i, X_{i+1}; \theta_0)|X_i] = \varpi$ or $E[g(X_i, X_{i+1}; \theta_0)|X_i] = O(\triangle)$, we can perform our following reasoning and calculation with only very slight modification in the finial step for working out the expectation.

This chapter is organized like this. Section 2.1 introduces some related works, concepts and theories. Section 2.2 presents our major theory and the theoretical bias and variance for O-U processes and CIR process. Section 2.3 includes some plots to help visualize the trend of the bias with respect to the parameter values. Simulation results, as well as theoretical results when value of parameters are given, are shown in Section 2.3. Section 2.5 constitutes all technical details.

2.1 Introduction

One of the most frequently used parameter estimation method in Statistics is the maximal likelihood estimation. But either because the full likelihood is not available especially for diffusion process, or the effect is not as good as desired, some alternative ones had been proposed. Among them, most well known ones are the Generalized Method of Moments proposed by Hansen and Scheinkman (1995), pseudo likelihood approach by Nowman (1997), the martingale estimating equation suggested by Bibby and Sørensen (1995), and the approximate likelihood method through expansions by Aït-Sahalia (2002).

More recently, Nkurunziza (2013) considered the general inference problem of the drift parameters matrices of m independent multivariate diffusion processes. Krumscheid (2013) et. al. proposed a novel algorithm for estimating both the drift and the diffusion coefficients in the effective dynamics based on a semi-parametric framework. Varughese (2013) proposed a cumulant truncation procedure to approximate MLE for parameter estimation for multivariate diffusion systems.

The following two papers are most closely related to our work.

In Tang and Chen (2009) [2] , the bias for one dimensional O-U process and CIR process are derived by taking expectation on the explicitly derived ML estimator or pseudo likelihood

[2] There is a typo in the result for Theorems 3.2.1. Based on their results about β in their working paper, the final result for bias of κ should be the same as in our result, which is $E(\hat{\kappa}) = \kappa + 4T^{-1} + \frac{2\kappa}{n} + o(n^{-1})$. Here we assume $o(n^{-1} + T^{-2}) = o(n^{-1})$.

estimator based on Nowman (1997) discretization. Their result are both for fixed sampling interval ($\triangle$ treated as a constant) or diminishing interval ($\triangle \to 0$, $n \to \infty$, while $n\triangle \to \infty$.).

Wang, Philips and Yu (2011) conducted the OLS estimation directly based on the exact discrete time model given by Phillips (1972) corresponding to

$$dX(t) = (A(\theta)X(t) + B(\theta))dt + dW(t), X(0) = X_0.$$

Their model is restricted to ensure stationary mean reverting and sampled at fixed time interval h and known B, yet their estimator for A adopts two approximated forms when $h \to 0$. [3]

We are inspired to derive the bias and variance forms of multi-dimensional stochastic differential equations under general estimating equation for diminishing intervals with exact original distribution.

In the following, we introduce the univariate and bivariate Ornstein-Uhlenbeck Processes, as well as the Cox-Ingersoll-Ross processes respectively. Some simple tensor notations are introduced too.

2.1.1 O-U Processes and CIR Processes

A one dimensional Ornstein-Uhlenbeck (O-U) diffusion process is a stochastic process which satisfies the following diffusion equation:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{1.2}$$

where $\{B_t\}_{t \geq 0}$ denotes the standard Brownian motion.

The two dimensional Ornstein-Uhlenbeck diffusion process is

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{1.3}$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} B_{1t} \\ B_{2t} \end{pmatrix}, \tag{1.4}$$

[3] Hence, the CIR process is estimated with the the pseudo likelihood under modified dynamics: the process when the diffusion part is treated as piece-wisely constant.

where $\kappa_{2\times 1}$, α_2 and $\sigma_{2\times 2}$ are constant vectors or matrix, and $\begin{pmatrix} B_{1t} \\ B_{2t} \end{pmatrix}_{t\geq 0}$ is a either a standard two dimensional Brownian motion or a two dimensional correlated Brownian motion.

The O-U process is stationary, Gaussian, and Markovian. It is slightly more complex than pure Brownian motion with one more drift term. The process tends to drift towards its long-term mean or mean reversion level, α, after extended time. This property is referred to as mean-reverting. κ is called mean reversion rate and σ is the volatility parameter. These unique properties enable O-U process to become one of the most popular models in the financial world. A one dimensional O-U process is called Vasicek model when used to mimic the evolution of interest rates. Please refer to Vasicek (1977) for more details.

Under the O-U process, the conditional transitional density and the marginal distribution, can be explicitly written out. Please refer to the technical proofs part in Section 2.5.2 and Section for details. We can calculate the corresponding conditional expectation of the moments under these known distributions. Or we can work it out through the sum of a series of infinitesimal generators apply to these moment functions based on our convergence result under some regularity conditions. The two results will be approximately identical. This justifies our introduction of the later methods, which will enable our estimation method applicable to cases when transitional density is not available. Example of how to apply our generator method for calculating the conditional expectation is provided in equation (5.14) in section 2.5.2.

A one dimensional CIR process is with the form

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t.$$

where $\{B_t\}_{t\geq 0}$ denotes the standard Brownian motion. The two dimensional CIR process is

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t,$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}\begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix}dt + \begin{pmatrix} \sigma_{11}\sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{X_{2t}} \end{pmatrix} d\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix}.$$

Here, $\{B_t\}_{t\geq 0}$ is a two dimensional standard Brownian. $\kappa_{2\times 1}$ and α_2 are two real vectors. $\sigma_{2\times 2}$ is a constant matrix with two zero components under our assumption.

The CIR process and the O-U process are defined very similarly, only O-U with X_t in the diffusion coefficient, while CIR is defined with $\sqrt{X_t}$. But their distribution and hence properties are quite different. For example, the one dimensional CIR process is no longer normal, but a non-central chi-square distribution, with stationary distribution gamma distributed. Please refer to Cox, Ingersoll (1985) and Ross for more details.

For more information about stochastic diffusion processes, please refer to Karatzas and Shreve (1988).

2.1.2 Tensor Notations

Let us introduce the tensor notations. Please refer to McCullagh (1987) for details.

Denote

$$f(X_i, X_{i+1}; \eta) = (f^1(X_i, X_{i+1}; \eta), \ldots, f^r(X_i, X_{i+1}; \eta)).$$

Here, $f^j(X_i, X_{i+1}; \eta)$ is the j-th component of $f(X_i, X_{i+1}; \eta)$. Likewise, η^j is the j-th component of η. We employ an convention in the tensor method that repeated indices are summed over, namely

$$u_{jm} g^m(X_\alpha, X_{\alpha+1}; \theta_0) = \sum_m u_{jm} g^m(X_\alpha, X_{\alpha+1}; \theta_0).$$

We need the following tenosor notations in our analysis:

$$\begin{aligned}
\beta^{j;j_1,\ldots,j_k}(X_i, X_{i+1}) &:= E[\frac{\partial^k f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^{j_1}, \ldots, \partial \eta^{j_k}}], \\
b_i^{j;j_1,\ldots,j_k}(X_i, X_{i+1}) &:= \frac{\partial^k f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^{j_1}, \ldots, \partial \eta^{j_k}} - \beta^{j;j_1,\ldots,j_k}, \\
B^{j;j_1,\ldots,j_k}(X_i, X_{i+1}) &:= \frac{1}{n} \sum_{i=1}^{n} b_i^{j;j_1,\ldots,j_k}(X_i, X_{i+1}),
\end{aligned}$$

$$\gamma^{b_1(j;j_1,\ldots,j_m;\ldots,k;k_1,\ldots,k_n);\ldots,b_s(i;i_1,\ldots,i_p;,\ldots;l;l_1,\ldots,l_q)} :=$$
$$E[A^{b_1}(\eta_0)\{(\frac{\partial^m f^j(X_\alpha,X_{\alpha+1};\theta_0)}{\partial\eta^{j_1},\ldots,\partial\eta^{j_m}}\times\cdots\times\frac{\partial^n f^k(X_\alpha,X_{\alpha+1};\theta_0)}{\partial\eta^{k_1},\ldots,\partial\eta^{k_n}})$$
$$\times\{\cdots\times\{A^{b_s}(\eta_0)(\frac{\partial^p f^i(X_\alpha,X_{\alpha+1};\theta_0)}{\partial\eta^{i_1},\ldots,\partial\eta^{i_p}}\times\frac{\partial^q f^l(X_\alpha,X_{\alpha+1};\theta_0)}{\partial\eta^{l_1},\ldots,\partial\eta^{l_q}})\}\}\}],$$
$$\zeta^j := \frac{1}{6}\sum_{k,l.m}\int_{\eta_0^k}^{\eta^k}\int_{\eta_0^l}^{\eta^l}\int_{\eta_0^m}^{\eta^m}\int_{\eta_0^n}^{\eta^n}\frac{\partial^3 f^j(t)}{\partial\eta^m\partial\eta^l\partial\eta^k}(\eta^k-t_1)(\eta^l-t_2)(\eta^m-t_3)dt_1dt_2dt_3,$$
$$\gamma^{b(j,k)} = E[A^b(\theta_0)\{u_{jm}g^m(X_\alpha,X_{\alpha+1};\theta_0)\cdot u_{kn}g_n(X_\alpha,X_{\alpha+1};\theta_0)\}],$$
$$\gamma^{b(j,k;k)} = E[A^b(\theta_0)\{\frac{\partial f^j(X_\alpha,X_{\alpha+1};\theta_0)}{\partial\eta^k}\cdot f^k(X_\alpha,X_{\alpha+1};\theta_0)\}]$$
$$= E[A^b(\theta_0)\{\frac{\partial(\sum_{m=1}^p u_{jm}g_m(X_\alpha,X_{\alpha+1};\theta_0))}{\partial\theta_n}\cdot v_{kn}\cdot\sum_{n=1} pu_{kn}g_n(X_\alpha,X_{\alpha+1};\theta_0)\}].$$

Furthermore, using the infinitesimal generator,

$$\begin{aligned} E[f^j(X_i,X_{i+1};\eta_0)|X_i=x_0] &= \sum_{k=0}^{K}A^k(\eta_0)f^j(x_0,x_0;\eta_0)\frac{\triangle^k}{k!} \\ &\quad +E[A^{K+1}(\eta_0)\cdot f^j(x_0,X_{i+\triangle^*};\eta_0)|X_i=x_0]\frac{\triangle^{K+1}}{(K+1)!}, \end{aligned} \tag{1.5}$$

where $0<\triangle^*<\triangle$.

2.2 Theory

In this section, we first present the theoretical forms of bias and variance for parameter estimators obtained via estimating parameters of an stochastic differential equation via the general estimating equation approach. Then, we list the theoretical forms of bias and variance for various O-U processes and CIR processes. Please refer to the Section 2.5 for technique details about how the theories are proved and how the results for those selected processes are derived.

2.2.1 Bias and Variance for General Diffusion Process

Now we want to explore the high order property of the estimator $\hat{\theta}$, which is gained , under this specification, so we make some further assumptions below, and mainly use the Taylor expansion, tensor method, properties of strong mixing processes to achieve our target.

Chen and Cui (2007) applied tensor method to explore the second order properties of empirical likelihood with moment restriction. Inspired by their proceedings, we assume

(A1) $rank(E[\frac{\partial}{\partial\theta}g(X_i,X_{i+1};\theta_0)]) = p$, and there exist non-singular matrices $U_{r\times r}$ and $V_{p\times p}$, such that

$$V^T \cdot E[\frac{\partial}{\partial\theta}g(X_i,X_{i+1};\theta_0)]\cdot U^T = \begin{pmatrix} I_{p\times p} \\ 0_{(r-p)\times p} \end{pmatrix},$$

where $I_{p\times p}$ is the $p-$dim identity matrix and $0_{(r-p)\times p}$ is the $(r-p)\times p$ null matrix.

We reparameterize $\eta_{p\times 1} := V^{-1}\theta$, and define

$$\begin{aligned} f(\eta)_{r\times 1} &:= f(X_i,X_{i+1};\eta) := U\cdot g(X_i,X_{i+1};\theta), \text{ and} && (2.6)\\ E[\frac{\partial}{\partial\eta}f(\eta_0)] &= E[\frac{\partial\theta}{\partial\eta}\cdot\frac{\partial f}{\partial\theta}] = E[\frac{\partial V\eta}{\partial\eta}\cdot\frac{\partial Ug}{\partial\theta}] \\ &= E[V^T\cdot\frac{\partial}{\partial\theta}g(X_i,X_{i+1};\theta_0)\cdot U^T] = \begin{pmatrix} I_{p\times p} \\ 0_{(r-p)\times p} \end{pmatrix}. && (2.7)\end{aligned}$$

Without loss of generality, we first consider the properties of the transformed parameter $\eta := V^{-1}\theta$, which can then translated to results regarding the original parameter θ.

The main result will gives out the form of the bias and variance for estimating the desired parameter. Before showing that, similar as in Sargan (1976), we employ the following theory to control the remainder terms and justify that they are ignorable in the tensor expansion.

Theorem 2.2.1 *Denote $\phi_T(\eta) := \frac{1}{n}\sum_{i=1}^{n} f(X_i,X_{i+1};\eta)$. Assume*

(i) There exist a hypersphere S and $N_0\in\mathbb{N}$, such that ϕ_T has uniformly bounded derivatives up to order r for all $p\times 1$ vector $\eta\in S$ and all $n\geq N_0$.

(ii) Let $\hat\eta$ be an estimator of η which has a probability density $f(\eta)$, with real value $\eta_0\in S$, such that

$$E||\hat\eta-\eta_0||^R = O(n^{-\gamma R}),$$

as $n\to\infty$, for some $\gamma\geq 0$ and $R\geq 1$.

(iii) For some $K \geq 1$, $E(|\phi_T(\hat{\eta})|^K) = O(n^{\lambda})$ *for some* $\lambda > 0$.

Then, for any $1 \leq k \leq \frac{K(R-r)}{R+\lambda/\gamma}$,

$$E(|\phi_T(\hat{\eta})|^k) - E(|\phi_{Tr}(\hat{\eta})|^k) = O(n^{-\gamma r}),$$

where $\phi_{Tr}(\hat{\eta}) := \sum_{s=0}^{r-1} \frac{1}{s!}[(\sum_{i=1}^{p}(\hat{\eta}_i - \eta_{0i})\frac{\partial}{\partial \eta_i})^s \phi_T(\hat{\eta})]|_{\hat{\eta}=\eta_0}$ *is the* $(r-1)-th$ *order Taylor Series expression about* $\eta = \eta_0$, r *is a positive integer less than or equal to* R, *and* $(r-1)k \leq R$.

To formally attain our result of the bias and variance, the following definitions and assumptions are introduced.

Let $(X_t, t \in \mathbb{Z})$ be a strictly stationary process, its strong mixing coefficient of order k is defined as

$$\alpha(k) = \sup_{B \in \sigma(X_s, s \leq t) C \in \sigma(X_s, s \geq t+k)} |P(B \bigcap C) - P(B)P(C)|, \;\; k \geq 1.$$

X is said to be strong mixing, or $\alpha-$mixing if $\lim_{k\to\infty} \alpha(k) = 0$. Please refer to Bosq (1998) for more details.

The three assumptions needed are:

A2 $\{X_t\}$ is strong mixing with mixing coefficient $\alpha(n)$ satisfying $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$, for some $\delta > 0$.

Then $b_j^{j;j_1,\ldots,j_k n}{}_{j=1}$ should be strong mixing too with coefficient $\beta(n)$ satisfying

$$\beta(n) \leq \alpha(n), \quad \mathcal{F}_{b_i} \subset \mathcal{F}_{X_i}.$$

A3 $E(b_i^{j;j_1,\ldots,j_k})^{2+\delta} < \infty$.

A4 $\sigma^2_{j;j_1,\ldots,j_k} := E(b_0^{j;j_1,\ldots,j_k})^2 + 2\sum_{i=1}^{\infty} E(b_0^{j;j_1,\ldots,j_k} b_i^{j;j_1,\ldots,j_k}) \neq 0$.

We also need the following convergence theory from Bosq (1998, p36) in the process of our proof.

Theorem 2.2.2 *Suppose that* $(X_t, t \in \mathbb{Z})$ *is a zero-mean real-valued strictly stationary process such that for some* $\gamma > 2$ *and some* $\delta > 0$

$$E|X_t|^{\gamma} < \infty, \text{ and}$$

$$\alpha(k) \leq ak^{-\beta}$$

where a is a positive constant and $\beta > \frac{\gamma}{\gamma-2}$, then if $\sigma^2 = \sum_{k=-\infty}^{\infty} Cov(X_0, X_k) > 0$ we have

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{w} N \sim \mathcal{N}(0,1).$$

Based on the above assumptions and theories, our main result can be established. It is summarized in Theorem 2.2.3. Please refer to the technique proof in Section 2.5 for details.

Theorem 2.2.3 *For parameters in stochastic differential equation*

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dB_t,$$

the bias and the variance of the estimator $\hat{\theta}_m$ are

$$bias(\hat{\theta}_m) = \frac{1}{n}\sum_{j=1}^{p}\sum_{b=0}^{B}\frac{\triangle^b}{b!}v_{mj}(\gamma^{b(j,k;k)} - \frac{1}{2}\beta^{j,kl}\gamma^{b(k;l)}) + o(\triangle^b),$$

$$var(\hat{\theta}_m) = \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{b=0}^{B}\frac{1}{n}\frac{\triangle^b}{b!}v_{mj}v_{mk}\gamma^{b(j;k)}.$$

2.2.2 Bias and Variance in One Dimensional O-U Process

Tang and Chen (2009) has already given the bias and variance forms for one dimensional O-U process. In this subsection, we apply our theory 2.2.3 to the special case of one dimensional O-U process to double check the correctness of our method. The relative simplicity of this process will also help us clearly demonstrate how our method is different from others' and how to be carried out.

We proceed following our new theorem below. Please refer to the technique proof part at the end of this chapter for more details.

For the one dimensional Ornstein-Uhlenbeck diffusion process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t,$$

where $\{B_t\}_{t\geq 0}$ denotes the standard Brownian motion.

By applying the Itô formula,

$$d(e^{\kappa t}X_t) = \kappa e^{\kappa t}X_t dt + e^{\kappa t}dX_t = e^{\kappa t}\kappa\alpha dt + e^{\kappa t}\sigma dW_t,$$

$$X_t = e^{-\kappa t}X_0 + \int_0^t e^{\kappa(s-t)}\kappa\alpha ds + \int_0^t e^{\kappa(s-t)}\sigma dW_s.$$

Hence, the conditional distribution is

$$X_i|X_{i-1} \sim N(X_{i-1}e^{-\kappa\triangle} + \alpha(1-e^{\kappa\triangle}), \frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\triangle})),$$

and the stationary distribution is $N(\alpha, \frac{1}{2}\sigma^2\kappa^{-1})$.

Then the likelihood function is

$$L := \sum_{i=1}^{n}[\frac{1}{2}\log\kappa - \frac{1}{2}\log\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log(1-e^{-2\kappa\triangle}) - \frac{(X_i-\alpha-e^{-\kappa\triangle(X_{i-1}-\alpha)})^2}{\sigma^2\kappa^{-1}(1-e^{-2\kappa\triangle})}].$$

We adopt the estimating equation

$$\begin{cases} g^1 := \frac{\partial L}{\partial\kappa} \\ g^2 := \frac{\partial L}{\partial\alpha} \\ g^3 := \frac{\partial L}{\partial\sigma^2} \end{cases}.$$

Then we can work out $V = I$,

$$U = \begin{pmatrix} -\frac{1-e^{-2\kappa\triangle}}{\triangle^2 e^{-2\kappa\triangle}} & 0 & -\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2\kappa\triangle}} + \frac{2\sigma^2}{\triangle} \\ 0 & -\frac{\sigma^2(1+e^{-\kappa\triangle})}{2\kappa(1-e^{-\kappa\triangle})} & 0 \\ -\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2\kappa\triangle}} + \frac{2\sigma^2}{\triangle} & 0 & -\frac{\sigma^4(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} + \frac{4\sigma 4}{\kappa\triangle} - \frac{2\sigma^4(1+e^{-2\kappa\triangle})}{1-e^{-2\kappa\triangle}} \end{pmatrix}.$$

So

$$
\begin{aligned}
& n \cdot bias(\kappa) \\
= \; & n \cdot bias(\eta^1) \\
= \; & nE[B^{1,k}B^k] - \frac{n}{2}\beta^{1,kl}E[B^kB^l] \\
= \; & E[\frac{\partial f^1(X_i, X_{i+1}; \eta_0)}{\partial \eta^k} f^k(X_i, X_{i+1}; \eta_0)] + \frac{1}{n}\sum_{i=1}^{n-1}(n-i)E[\frac{\partial f^1(X_{i+1}, X_{i+2}; \eta_0)}{\partial \eta^k} f^k(X_1, X_2; \eta_0)] \\
& -\frac{1}{2}E[\frac{\partial^2 f^1(X_1, X_2; \eta_0)}{\partial \eta^k \partial \eta^l}]E[f^k(X_1, X_2; \eta_0) f^l(X_1, X_2; \eta_0)] \\
= \; & E[u_{1i}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{kj} g^j(X_\alpha, X_{\alpha+1}; \theta_0)] \\
& -\frac{1}{2}E[u_{1i}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k \partial \theta^l}]E[u_{ki} g^i(X_\alpha, X_{\alpha+1}; \theta_0) \cdot u_{lj} g^j(X_\alpha, X_{\alpha+1}; \theta_0)] \\
& +\sum_{\alpha=1}^{n-1}\frac{n-\alpha}{n}E[u_{1i}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{kj} g^j(X_1, X_2; \theta_0)] \\
= \; & \frac{4}{\triangle} + 2\kappa + o(1),
\end{aligned}
$$

i.e. $bias(\hat{\kappa}) = \frac{4}{T} + \frac{2\kappa}{n} + o(\frac{1}{n})$.

Similarly, we can work out

$$
\begin{aligned}
bias(\hat{\sigma^2}) &= -\frac{2\sigma^2}{n} + o(\frac{1}{n}), \\
bias(\hat{\alpha}) &= o(\frac{1}{n}),
\end{aligned}
$$

$$
\begin{aligned}
var(\hat{\kappa}) &= \frac{2\kappa}{T} + o(\frac{1}{T}), \\
var(\hat{\sigma^2}) &= \frac{2\sigma^4}{n} + o(\frac{1}{n}), \\
var(\hat{\alpha}) &= \frac{\sigma^2}{T\kappa^2} + o(\frac{1}{T}),
\end{aligned}
$$

$$
\begin{aligned}
cov(\hat{\kappa}, \hat{\sigma}^2) &= o(\frac{1}{n}), \\
cov(\hat{\kappa}, \hat{\alpha}) &= \frac{2\sigma^2\kappa}{n}, \\
cov(\hat{\sigma}^2, \hat{\alpha}) &= o(\frac{1}{n}).
\end{aligned}
$$

This finding is consistent with Tang and Chen (2009). Note that since the transitional density of Ornstein-Uhlenbeck diffusion process is fully known, we have employed both its known transitional density and its infinitesimal generator to calculate the expectations and the results are the same. Also, the way to arrive at the result of the bias and variance is by plugging in the specified drift and diffusion functions and applying corresponding conditional distribution to the forms of bais and variance, which are obtained by tensor method to general estimating equation. Please refer to the appendix for technique details.

2.2.3 Bias and Variance in Two Dimensional O-U Process Driven by Independent Brownian Motions

To see how our general theory works on multi-dimensional stochastic differential equations, we first apply it to two dimensional O-U process driven by independent Brownian motions.

We adopt the two dimensional Ornstein-Uhlenbeck diffusion process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t,$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix},$$

where $\kappa_{21} \neq 0$ (or it is a degenerated case) and $\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix}_{t\geq 0}$ is a standard two dimensional Brownian motion.

Because

$$e^{\kappa t}dX_t = e^{\kappa t}\kappa(\alpha - X_t)dt + e^{\kappa t}\sigma dW_t,$$

$$d(e^{\kappa t}X_t) = \kappa e^{\kappa t}X_t dt + e^{\kappa t}dX_t = e^{\kappa t}\kappa\alpha dt + e^{\kappa t}\sigma dW_t,$$

$$e^{\kappa t}X_t - X_0 = \int_0^t e^{\kappa s}\kappa\alpha ds + \int_0^t e^{\kappa s}\sigma dW_s,$$

we see,

$$X_t \quad = \quad e^{-\kappa\triangle}X_{t-1}+(I-e^{-\kappa\triangle})\alpha+e^{-\kappa t\triangle}\int_{(t-1)\triangle}^{t\triangle}e^{\kappa s}\sigma dW_s.$$

When $\kappa_{11}\neq\kappa_{22}$, let $b=\frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}$, we have [4]

$$\exp\begin{pmatrix}\kappa_{11} & 0\\ \kappa_{21} & \kappa_{22}\end{pmatrix}=\exp\left\{\begin{pmatrix}1 & 0\\ -b & 1\end{pmatrix}\begin{pmatrix}\kappa_{11} & 0\\ 0 & \kappa_{22}\end{pmatrix}\begin{pmatrix}1 & 0\\ b & 1\end{pmatrix}\right\}$$

$$= \begin{pmatrix}1 & 0\\ -b & 1\end{pmatrix}\exp\begin{pmatrix}\kappa_{11}t & 0\\ 0 & \kappa_{22}t\end{pmatrix}\begin{pmatrix}1 & 0\\ b & 1\end{pmatrix}$$

$$= \begin{pmatrix}1 & 0\\ -b & 1\end{pmatrix}\begin{pmatrix}e^{\kappa_{11}} & 0\\ 0 & e^{\kappa_{22}}\end{pmatrix}\begin{pmatrix}1 & 0\\ b & 1\end{pmatrix}=\begin{pmatrix}e^{\kappa_{11}} & 0\\ b(e^{\kappa_{22}}-e^{\kappa_{11}}) & e^{\kappa_{22}}\end{pmatrix}.$$

When $\kappa_{11}=\kappa_{22}$,

$$\exp\begin{pmatrix}\kappa_{11} & 0\\ \kappa_{21} & \kappa_{22}\end{pmatrix}=\exp\left\{\kappa_{11}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\kappa_{21}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}\right\}$$

$$= \exp\left\{\kappa_{11}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right\}\exp\left\{\kappa_{21}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}\right\}$$

$$= \begin{pmatrix}e^{\kappa_{11}} & 0\\ 0 & e^{\kappa_{22}}\end{pmatrix}\begin{pmatrix}1 & 0\\ \kappa_{21} & 1\end{pmatrix}=\begin{pmatrix}e^{\kappa_{11}} & 0\\ \kappa_{21}e^{\kappa_{11}} & e^{\kappa_{11}}\end{pmatrix}.$$

Let

$$U_i^0=e^{-\kappa\triangle}X_{i-1}+(I-e^{-\kappa\triangle})\alpha.$$

---

[4] So $\exp\left\{-\begin{pmatrix}\kappa_{11} & 0\\ \kappa_{21} & \kappa_{22}\end{pmatrix}\triangle\right\}=\begin{pmatrix}e^{-\kappa_{11}\triangle} & 0\\ b(e^{-\kappa_{22}\triangle}-e^{-\kappa_{11}\triangle}) & e^{-\kappa_{22}\triangle}\end{pmatrix}.$

We have when $\kappa_{11} \neq \kappa_{22}$,

$$
\begin{aligned}
U_i &= X_i - U_i^0 = \begin{pmatrix} X_{i,1} - U_{i1}^0 \\ X_{i,2} - U_{i2}^0 \end{pmatrix}, \\
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1), \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{22}\triangle + \frac{\kappa_{22}^2}{2}\triangle^2 - \frac{\kappa_{22}^3}{6}\triangle^3 + \frac{\kappa_{22}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2), \\
&\quad -\kappa_{21}\triangle\{1 - \frac{1}{2}\triangle(\kappa_{11} + \kappa_{22}) + \frac{1}{6}\triangle^2(\kappa_{11}^2 + \kappa_{11}\kappa_{22} + \kappa_{22}^2) \\
&\quad -\frac{1}{24}\triangle^3(\kappa_{11}^2 + \kappa_{22}^2)(\kappa_{11} + \kappa_{22})\}(X_{i-1,1} - \alpha_1).
\end{aligned}
$$

When $\kappa_{11} = \kappa_{22}$,

$$
\begin{aligned}
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1), \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2) \\
&\quad -\kappa_{21}\triangle(1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3)(X_{i-1,1} - \alpha_1).
\end{aligned}
$$

Denote

$$
V = var(e^{-\kappa t\triangle}\int_{(t-1)\triangle}^{t\triangle} e^{\kappa s}\sigma dW_s) = \int_{(t-1)\triangle}^{t\triangle} e^{\kappa(-t\triangle+s)}\sigma\sigma^T(e^{\kappa(-t\triangle+s)})^T ds.
$$

Then we find when $\kappa_{11} \neq \kappa_{22}$,

$$
\begin{aligned}
V &= \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}, \\
v_{11} &= \sigma_{11}^2\frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}} \\
&= \sigma_{11}^2\triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5), \\
v_{12} &= v_{21} = \frac{\kappa_{21}}{\kappa_{22} - \kappa_{11}}\sigma_{11}^2(\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} - \frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}}) \\
&= \sigma_{11}^2\kappa_{21}\triangle^2\{-\frac{1}{2} + \triangle(\frac{1}{2}\kappa_{11} + \frac{1}{6}\kappa_{22}) - \triangle^2(\frac{7}{24}\kappa_{11}^2 + \frac{1}{6}\kappa_{11}\kappa_{22} + \frac{1}{24}\kappa_{22}^2)\} + O(\triangle^5), \\
v_{22} &= (\frac{\kappa_{21}}{\kappa_{22} - \kappa_{11}})^2\sigma_{11}^2\{\frac{1 - e^{-2\kappa_{11}}}{2\kappa_{11}} - 2\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} + \frac{1 - e^{-2\kappa_{22}}}{2\kappa_{22}}\} + \sigma_{22}^2\frac{1 - e^{-2\kappa_{22}}}{2\kappa_{22}} \\
&= \sigma_{11}^2\kappa_{21}^2\triangle^3\{\frac{1}{3} - \triangle\frac{1}{4}(\kappa_{11} + \kappa_{22})\} + \sigma_{22}^2\triangle(1 - \kappa_{22}\triangle + \frac{2}{3}\kappa_{22}^2\triangle^2 - \frac{1}{3}\kappa_{22}^3\triangle^3) + O(\triangle^5).
\end{aligned}
$$

When $\kappa_{11} = \kappa_{22}$,

$$
\begin{aligned}
v_{11} &= \sigma_{11}^2 \triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5), \\
v_{12} &= v_{21} = \sigma_{11}^2\kappa_{21}\triangle^2(-\frac{1}{2} + \triangle\frac{2}{3}\kappa_{11} - \triangle^2\frac{1}{2}\kappa_{11}^2) + O(\triangle^5), \\
v_{22} &= \sigma_{11}^2\kappa_{21}^2\triangle^3(\frac{1}{3} - \frac{1}{2}\kappa_{11}\triangle) + \sigma_{22}^2\triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5).
\end{aligned}
$$

That is the main orders for the conditional expectation and variance of X_i given X_{i-1} when $\kappa_{11} = \kappa_{22}$ is the same as letting $\kappa_{11} = \kappa_{22}$ in the $\kappa_{11} \neq \kappa_{22}$ case. Therefore, we can only derive those results for $\kappa_{11} \neq \kappa_{22}$ and then let $\kappa_{11} = \kappa_{22}$ to get the result for $\kappa_{11} = \kappa_{22}$.

That is we know the transitional distribution

$$
X_t | X_{t-1} \sim N_2(U_i^0, V),
$$

and the stationary distribution

$$
X_t \sim N_2(\alpha, V_0),
$$

where $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ and $V_0 = \begin{pmatrix} \frac{\sigma_{11}^2}{2\kappa_{11}} & b\sigma_{11}^2(\frac{1}{\kappa_{11}+\kappa_{22}} - \frac{1}{2\kappa_{11}}) \\ b\sigma_{11}^2(\frac{1}{\kappa_{11}+\kappa_{22}} - \frac{1}{2\kappa_{11}}) & b^2\sigma_{11}^2(\frac{1}{2\kappa_{11}} - \frac{2}{\kappa_{11}+\kappa_{22}} + \frac{1}{2\kappa_{22}}) + \frac{\sigma_{22}^2}{2\kappa_{22}} \end{pmatrix}$.

Then the log likelihood function is

$$
l_n(\theta) = \sum_{i=1}^{n} \{-\log(2\pi) - \frac{1}{2} l^*(\theta)\},
$$

where $\theta = \{\alpha_1, \alpha_2, \kappa_{11}, \kappa_{22}, \kappa_{21}, \sigma_{11}^2, \sigma_{22}^2\}$. Let

$$
l_n^*(\theta) = \log|V| + U_i^T V^{-1} U_i.
$$

And we adopt the estimating equation to be

$$
\frac{\partial l_n^*(\theta)}{\partial \theta} = 0.
$$

Based on our main result, in Theorem 2.2.3 by plug in the specific form of the drift function and the diffusion function, calculate the expectations either through the infinitesimal generator or the known transitional density of two dimensional O-U process, we find that the biases are:

$$
\begin{aligned}
bias(\hat{\alpha}_1) &= o(\frac{1}{n}),\\
bias(\hat{\alpha}_1) &= o(\frac{1}{n}),\\
bias(\hat{\kappa}_{11}) &= \frac{4+2\Delta\kappa_{11}}{n\Delta} + o(\frac{1}{n}),\\
bias(\hat{\kappa}_{22}) &= \frac{(\kappa_{11}+\kappa_{22})\,\sigma_{22}^2}{n\Delta\left(\kappa_{21}^2\sigma_{11}^2+(\kappa_{11}+\kappa_{22})^2\sigma_{22}^2\right)^2}\left(4\kappa_{11}\kappa_{21}^2\sigma_{11}^2+6\kappa_{21}^2\kappa_{22}\sigma_{11}^2+4\kappa_{11}^3\sigma_{22}^2+14\kappa_{11}^2\kappa_{22}\sigma_{22}^2+\right.\\
&\quad \left.16\kappa_{11}\kappa_{22}^2\sigma_{22}^2+6\kappa_{22}^3\sigma_{22}^2\right)\\
&\quad +\frac{(\kappa_{11}+\kappa_{22})\,\sigma_{22}^2}{n\left(\kappa_{21}^2\sigma_{11}^2+(\kappa_{11}+\kappa_{22})^2\sigma_{22}^2\right)^2}\left(-\kappa_{11}\kappa_{21}^2\kappa_{22}\sigma_{11}^2+\kappa_{21}^2\kappa_{22}^2\sigma_{11}^2+2\kappa_{11}^3\kappa_{22}\sigma_{22}^2+7\kappa_{11}^2\kappa_{22}^2\sigma_{22}^2\right.\\
&\quad \left.+8\kappa_{11}\kappa_{22}^3\sigma_{22}^2+3\kappa_{22}^4\sigma_{22}^2\right)+o(\frac{1}{n}),\\
bias(\hat{\kappa}_{21}) &= \frac{\kappa_{21}\sigma_{22}^2\left(6\kappa_{21}^2\kappa_{22}\sigma_{11}^2+4\kappa_{11}^3\sigma_{22}^2+14\kappa_{11}^2\kappa_{22}\sigma_{22}^2+6\kappa_{22}^3\sigma_{22}^2+4\kappa_{11}\left(\kappa_{21}^2\sigma_{11}^2+4\kappa_{22}^2\sigma_{22}^2\right)\right)}{n\Delta\left(\kappa_{21}^2\sigma_{11}^2+(\kappa_{11}+\kappa_{22})^2\sigma_{22}^2\right)^2}\\
&\quad +\frac{\kappa_{21}\sigma_{22}^2}{n\left(\kappa_{21}^2\sigma_{11}^2+(\kappa_{11}+\kappa_{22})^2\sigma_{22}^2\right)^2}(\kappa_{11}+\kappa_{22})\left(2\kappa_{11}\kappa_{21}^2\sigma_{11}^2+\kappa_{21}^2\kappa_{22}\sigma_{11}^2+2\kappa_{11}^3\sigma_{22}^2\right.\\
&\quad \left.+8\kappa_{11}^2\kappa_{22}\sigma_{22}^2+9\kappa_{11}\kappa_{22}^2\sigma_{22}^2+3\kappa_{22}^3\sigma_{22}^2\right)+o(\frac{1}{n}),\\
bias(\hat{\sigma}_{11}^2) &= \frac{2\sigma_{11}^2}{n}+o(\frac{1}{n}),\\
bias(\hat{\sigma}_{22}^2) &= \frac{\sigma_{22}^2\left(-18\kappa_{21}^2\sigma_{11}^2+6\left(\kappa_{11}^2+4\kappa_{11}\kappa_{22}+3\kappa_{22}^2\right)\sigma_{22}^2\right)}{6n\left(\kappa_{21}^2\sigma_{11}^2+(\kappa_{11}+\kappa_{22})^2\sigma_{22}^2\right)}+o(\frac{1}{n}).
\end{aligned}
$$

And the variances and covariances are:

$$
\begin{aligned}
var(\hat{\alpha}_1) &= \frac{\sigma_{11}^2}{n\Delta\kappa_{11}^2}+o(\frac{1}{n\Delta}),\\
var(\hat{\alpha}_2) &= \frac{\frac{\kappa_{21}^2\sigma_{11}^2}{\kappa_{11}^2}+\sigma_{22}^2}{n\Delta\kappa_{22}^2}+o(\frac{1}{n\Delta}),
\end{aligned}
$$

$$\begin{aligned}
var(\hat{\kappa}_{11}) &= \frac{2\kappa_{11}}{n\Delta} + \frac{2\kappa_{11}^2}{n} + o(\frac{1}{n}), \\
var(\hat{\kappa}_{22}) &= \frac{2\kappa_{22}\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2}{n\Delta\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)} + \frac{\kappa_{22}^2\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\left(\kappa_{21}^2\sigma_{11}^2+2\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)}{n\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)^2} \\
&\quad + o(\frac{1}{n}), \\
var(\hat{\kappa}_{21}) &= \frac{2\left(\kappa_{11}+\kappa_{22}\right)\sigma_{22}^2\left(\kappa_{21}^2\sigma_{11}^2+\kappa_{11}\left(\kappa_{11}+\kappa_{22}\right)\sigma_{22}^2\right)}{n\Delta\sigma_{11}^2\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)} \\
&\quad + \frac{\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\left(\kappa_{21}^4\sigma_{11}^4+2\kappa_{21}^2\left(\kappa_{11}^2+\kappa_{11}\kappa_{22}+\kappa_{22}^2\right)\sigma_{11}^2\sigma_{22}^2+\kappa_{11}^2\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^4\right)}{n\sigma_{11}^2\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)^2} \\
&\quad + o(\frac{1}{n}),
\end{aligned}$$

$$\begin{aligned}
var(\hat{\sigma}_{11}^2) &= \frac{2\sigma_{11}^4}{n} + o(\frac{1}{n}), \\
var(\hat{\sigma}_{22}^2) &= \frac{2\sigma_{22}^4}{n} + o(\frac{1}{n}),
\end{aligned}$$

$$\begin{aligned}
cov(\hat{\alpha}_1,\hat{\alpha}_2) &= -\frac{\kappa_{21}\sigma_{11}^2}{n\Delta\kappa_{11}^2\kappa_{22}} + o(\frac{1}{n\Delta}), \\
cov(\hat{\kappa}_{22},\hat{\kappa}_{21}) &= \frac{2\kappa_{21}\kappa_{22}\left(\kappa_{11}+\kappa_{22}\right)\sigma_{22}^2}{n\Delta\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)} \\
&\quad + \frac{\kappa_{21}\kappa_{22}\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}^2+3\kappa_{11}\kappa_{22}+2\kappa_{22}^2\right)\sigma_{22}^2\right)}{n\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)^2} + o(\frac{1}{n}), \\
cov(\hat{\kappa}_{11},\hat{\sigma}_{11}^2) &= \frac{2\kappa_{11}\sigma_{11}^2}{n} + o(\frac{1}{n}), \\
cov(\hat{\kappa}_{22},\hat{\sigma}_{22}^2) &= \frac{2\kappa_{22}\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^4}{n\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)} + o(\frac{1}{n}), \\
cov(\hat{\kappa}_{21},\hat{\sigma}_{22}^2) &= \frac{2\kappa_{21}\kappa_{22}\left(\kappa_{11}+\kappa_{22}\right)\sigma_{22}^4}{n\left(\kappa_{21}^2\sigma_{11}^2+\left(\kappa_{11}+\kappa_{22}\right)^2\sigma_{22}^2\right)} + o(\frac{1}{n}),
\end{aligned}$$

while $cov(\hat{\kappa}_{11},\hat{\kappa}_{22})$, $cov(\hat{\kappa}_{11},\hat{\kappa}_{21})$, $cov(\hat{\kappa}_{11},\hat{\sigma}_{22}^2)$, $cov(\hat{\kappa}_{22},\hat{\sigma}_{11}^2)$, $cov(\hat{\kappa}_{21},\hat{\sigma}_{11}^2)$, and $cov(\hat{\sigma}_{11}^2,\hat{\sigma}_{22}^2)$ are all $o(\frac{1}{n})$.

We can see the order of the two dimensional case is largely the same as the one dimensional case.

Above results for two dimensional O-U process are partially harvested with the aid of Mathematica due to the enormous amount of calculation involved.

2.2.4 Bias and Variance in Two Dimensional O-U Process Driven by Correlated Brownian Motions

We adopt the two dimensional Ornstein-Uhlenbeck diffusion process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t,$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}\begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} B_{1t} \\ B_{2t} \end{pmatrix},$$

where $\kappa_{21} \neq 0$ (or it is a degenerated case) and $\begin{pmatrix} B_{1t} \\ B_{2t} \end{pmatrix}_{t\geq 0}$ is a standard two dimensional Brownian motion with $corr(dB_{1t}, dB_{2t}) = \rho$.

Or equivalently,

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t,$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}\begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{11}\rho & \sigma_{22}\sqrt{1-\rho^2} \end{pmatrix} d\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix},$$

where $\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix}_{t\geq 0}$ is a standard two dimensional Brownian motion with $corr(dW_{1t}, dW_{2t}) = 0$.

Let

$$U_i^0 = e^{-\kappa\triangle}X_{i-1} + (I - e^{-\kappa\triangle})\alpha.$$

We have when $\kappa_{11} \neq \kappa_{22}$,

$$
\begin{aligned}
U_i &= X_i - U_i^0 = \begin{pmatrix} X_{i,1} - U_{i1}^0 \\ X_{i,2} - U_{i2}^0 \end{pmatrix}, \\
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1), \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{22}\triangle + \frac{\kappa_{22}^2}{2}\triangle^2 - \frac{\kappa_{22}^3}{6}\triangle^3 + \frac{\kappa_{22}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2), \\
&\quad -\kappa_{21}\triangle\{1 - \frac{1}{2}\triangle(\kappa_{11} + \kappa_{22}) + \frac{1}{6}\triangle^2(\kappa_{11}^2 + \kappa_{11}\kappa_{22} + \kappa_{22}^2) \\
&\quad -\frac{1}{24}\triangle^3(\kappa_{11}^2 + \kappa_{22}^2)(\kappa_{11} + \kappa_{22})\}(X_{i-1,1} - \alpha_1).
\end{aligned}
$$

When $\kappa_{11} = \kappa_{22}$,

$$
\begin{aligned}
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1), \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2) \\
&\quad -\kappa_{21}\triangle(1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3)(X_{i-1,1} - \alpha_1).
\end{aligned}
$$

Denote

$$
V = var(e^{-\kappa t\triangle}\int_{(t-1)\triangle}^{t\triangle} e^{\kappa s}\sigma dW_s) = \int_{(t-1)\triangle}^{t\triangle} e^{\kappa(-t\triangle+s)}\sigma\sigma^T(e^{\kappa(-t\triangle+s)})^T ds.
$$

Then we find when $\kappa_{11} \neq \kappa_{22}$,

$$
V = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix},
$$

$$
\begin{aligned}
v_{11} &= \sigma_{11}^2 \frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}} \\
&= \sigma_{11}^2\triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5),
\end{aligned}
$$

$$
\begin{aligned}
v_{12} &= v_{21} = b\sigma_{11}^2(\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} - \frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}}) + \rho\sigma_{11}\sigma_{22}\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} \\
&= \sigma_{11}^2\kappa_{21}\triangle^2\{-\frac{1}{2} + \triangle(\frac{1}{2}\kappa_{11} + \frac{1}{6}\kappa_{22}) - \triangle^2(\frac{7}{24}\kappa_{11}^2 + \frac{1}{6}\kappa_{11}\kappa_{22} + \frac{1}{24}\kappa_{22}^2)\} \\
&\quad +\rho\sigma_{11}\sigma_{22}\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} + O(\triangle^5),
\end{aligned}
$$

$$
\begin{aligned}
v_{22} &= b^2\sigma_{11}^2\{\frac{1-e^{-2\kappa_{11}}}{2\kappa_{11}} - 2\frac{1-e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11}+\kappa_{22}} + \frac{1-e^{-2\kappa_{11}}}{2\kappa_{22}}\} + \sigma_{22}^2\frac{1-e^{-2\kappa_{22}}}{\kappa_{22}} \\
&\quad +\rho\sigma_{11}\sigma_{22}\frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}(\frac{1-e^{-2\kappa_{22}}}{2\kappa_{22}} - \frac{1-e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11}+\kappa_{22}}) \\
&= \sigma_{11}^2\kappa_{21}^2\triangle^3\{\frac{1}{3} - \triangle\frac{1}{4}(\kappa_{11}+\kappa_{22})\} + \sigma_{22}^2\triangle(1-\kappa_{22}\triangle + \frac{2}{3}\kappa_{22}^2\triangle^2 - \frac{1}{3}\kappa_{22}^3\triangle^3) \\
&\quad +\rho\sigma_{11}\sigma_{22}\frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}(\frac{1-e^{-2\kappa_{22}}}{2\kappa_{22}} - \frac{1-e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11}+\kappa_{22}}) + O(\triangle^5).
\end{aligned}
$$

When $\kappa_{11} = \kappa_{22}$,

$$
\begin{aligned}
v_{11} &= \sigma_{11}^2\triangle(1-\kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5), \\
v_{12} &= v_{21} = \sigma_{11}^2\kappa_{21}\triangle^2(-\frac{1}{2} + \triangle\frac{2}{3}\kappa_{11} - \triangle^2\frac{1}{2}\kappa_{11}^2) + \rho\sigma_{11}\sigma_{22}\frac{1-e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11}+\kappa_{22}} + O(\triangle^5), \\
v_{22} &= \sigma_{11}^2\kappa_{21}^2\triangle^3(\frac{1}{3} - \frac{1}{2}\kappa_{11}\triangle) + \sigma_{22}^2\triangle(1-\kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) \\
&\quad +\rho\sigma_{11}\sigma_{22}\frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}(\frac{1-e^{-2\kappa_{22}}}{2\kappa_{22}} - \frac{1-e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11}+\kappa_{22}}) + O(\triangle^5).
\end{aligned}
$$

The above analysis shows that the main orders for the conditional expectation and variance of X_i given X_{i-1} when $\kappa_{11} = \kappa_{22}$ is the same as letting $\kappa_{11} = \kappa_{22}$ in the $\kappa_{11} \neq \kappa_{22}$ case. Therefore, we can only derive those results for $\kappa_{11} \neq \kappa_{22}$ and then let $\kappa_{11} = \kappa_{22}$ to get the result for $\kappa_{11} = \kappa_{22}$.

That is we know the transitional distribution

$$X_t|X_{t-1} \sim N_2(U_i^0, V),$$

and the stationary distribution

$$X_t \sim N_2(\alpha, V_0),$$

where $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$, $V_0 = \begin{pmatrix} v_{011} & v_{012} \\ v_{021} & v_{022} \end{pmatrix}$, where

$$
\begin{aligned}
v_{011} &= \frac{\sigma_{11}^2}{2\kappa_{11}}, \\
v_{012} &= v_{021} = \frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}\sigma_{11}^2(\frac{1}{\kappa_{11}+\kappa_{22}} - \frac{1}{2\kappa_{11}}) + \rho\sigma_{11}\sigma_{22}\frac{1}{\kappa_{11}+\kappa_{22}}, \\
v_{022} &= (\frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}})^2\sigma_{11}^2(\frac{1}{2\kappa_{11}} - \frac{2}{\kappa_{11}+\kappa_{22}} + \frac{1}{2\kappa_{22}}) + \frac{\sigma_{22}^2}{2\kappa_{22}} + 2\rho\sigma_{11}\sigma_{22}\frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}(\frac{1}{2\kappa_{22}} - \frac{1}{\kappa_{11}+\kappa_{22}}).
\end{aligned}
$$

Then the the log likelihood function is

$$l_n(\theta) = \sum_{i=1}^{n} \{-\log(2\pi) - \frac{1}{2} l^*(\theta)\},$$

where $\theta = \{\alpha_1, \alpha_2, \kappa_{11}, \kappa_{22}, \kappa_{21}, \sigma_{11}^2, \sigma_{22}^2, \rho\}$,

$$l_n^*(\theta) = \log|V| + U_i^T V^{-1} U_i.$$

And we adopt the estimating equation to be

$$\frac{\partial l_n^*(\theta)}{\partial \theta} = 0.$$

Based on Theorem 2.2.3 and the above findings, we discover that the orders of bias for those interested parameters are of the same order as the independent case. The deduction process is aided with Mathematica. The middle processes and even the final deduced biases of those parameters takes large amount of pages to show. Hence, some simulation results are included in Section 2.3 to help with easy appreciation of the result.

2.2.5 Bias and Variance in Two Dimensional CIR Process

Consider the two dimensional CIR process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t,$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}\sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{X_{2t}} \end{pmatrix} d\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix}.$$

For the purpose of constructing likelihood function, we apply Nowman (1977)'s approximation method. For $s \in [(t-1)\triangle, t\triangle)$, define $Y_{1,s-1} = X_{(t-1)\triangle,1}$ and $Y_{2,s-1} = X_{(t-1)\triangle,2}$. The process used for likelihood derivation will be,

$$d\begin{pmatrix} X_{1s} \\ X_{2s} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_{1s} \\ \alpha_2 - X_{2s} \end{pmatrix} ds + \begin{pmatrix} \sigma_{11}\sqrt{Y_{1,s-1}} & 0 \\ 0 & \sigma_{22}\sqrt{Y_{2,s-1}} \end{pmatrix} d\begin{pmatrix} W_{1s} \\ W_{2s} \end{pmatrix}.$$

Write

$$
\begin{aligned}
U_i^0 &= e^{-\kappa\triangle}X_{i-1} + (I - e^{-\kappa\triangle})\alpha, \\
U_i &= X_i - U_i^0 = \begin{pmatrix} X_{i,1} - U_{i1}^0 \\ X_{i,2} - U_{i2}^0 \end{pmatrix}, \\
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1), \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{22}\triangle + \frac{\kappa_{22}^2}{2}\triangle^2 - \frac{\kappa_{22}^3}{6}\triangle^3 + \frac{\kappa_{22}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2), \\
&\quad -\kappa_{21}\triangle\{1 - \frac{1}{2}\triangle(\kappa_{11} + \kappa_{22}) + \frac{1}{6}\triangle^2(\kappa_{11}^2 + \kappa_{11}\kappa_{22} + \kappa_{22}^2) \\
&\quad -\frac{1}{24}\triangle^3(\kappa_{11}^2 + \kappa_{22}^2)(\kappa_{11} + \kappa_{22})\}(X_{i-1,1} - \alpha_1).
\end{aligned}
$$

Denote

$$
\begin{aligned}
V &= \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}, \\
v_{11} &= \sigma_{11}^2 Y_{1,t-1}\frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}} \\
&= \sigma_{11}^2 Y_{1,t-1}\triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5), \\
v_{12} &= v_{21} = \frac{\kappa_{21}}{\kappa_{22} - \kappa_{11}}\sigma_{11}^2 Y_{1,t-1}(\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} - \frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}}) \\
&= \sigma_{11}^2 Y_{1,t-1}\kappa_{21}\triangle^2\{-\frac{1}{2} + \triangle(\frac{1}{2}\kappa_{11} + \frac{1}{6}\kappa_{22}) - \triangle^2(\frac{7}{24}\kappa_{11}^2 + \frac{1}{6}\kappa_{11}\kappa_{22} + \frac{1}{24}\kappa_{22}^2)\} + O(\triangle^5), \\
v_{22} &= (\frac{\kappa_{21}}{\kappa_{22} - \kappa_{11}})^2\sigma_{11}^2 Y_{1,t-1}\{\frac{1 - e^{-2\kappa_{11}}}{2\kappa_{11}} - 2\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} + \frac{1 - e^{-2\kappa_{22}}}{2\kappa_{22}}\} + \sigma_{22}^2 Y_{2,t-1}\frac{1 - e^{-2\kappa_{22}}}{2\kappa_{22}} \\
&= \sigma_{11}^2 Y_{1,t-1}\kappa_{21}^2\triangle^3\{\frac{1}{3} - \triangle\frac{1}{4}(\kappa_{11} + \kappa_{22})\} + \sigma_{22}^2 Y_{2,t-1}\triangle(1 - \kappa_{22}\triangle + \frac{2}{3}\kappa_{22}^2\triangle^2 - \frac{1}{3}\kappa_{22}^3\triangle^3) + O(\triangle^5).
\end{aligned}
$$

Then the likelihood function is:

$$l_n(\theta) = \sum_{i=1}^{N}\{-log(1\pi) - \frac{1}{2}\log|V| - \frac{1}{2}U_i^T V^{-1} U_i\}.$$

Let
$$l_n^*(\theta) = \log|V| + U_i^T V^{-1} U_i.$$

Denote

$$U = (u_{ij}) = [(\frac{\partial^2 l^*}{\partial^2\theta})^T]^{-1},$$

$$g = (g^i) = \frac{\partial l^*}{\partial \theta}.$$

We need to derive:

$$\begin{aligned}
& n \cdot bias(\eta^j) \\
= \; & E[u_{ji}\frac{\partial g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial \theta^k} \cdot u_{ks}g^s(X_\alpha, X_{\alpha+1};\theta_0)] \\
& -\frac{1}{2}E[u_{ji}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial \theta^k \partial \theta^l}]\{E[u_{ks}g^s(X_\alpha, X_{\alpha+1};\theta_0) \cdot u_{lm}g^m(X_\alpha, X_{\alpha+1};\theta_0)] \\
& +\sum_{\alpha=j}^{n-1}\frac{n-\alpha}{n}E[u_{ji}\frac{\partial g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial \theta^k} \cdot u_{ks}g^s(X_1, X_2;\theta_0)] - E[u_{kl}g^l(X_\alpha, X_{\alpha+1};\theta_0)]E[u_{jm}\frac{g^m(X_\alpha, X_{\alpha+1};\theta_0)}{\partial \theta^k}].
\end{aligned}$$

We note that the CIR process implies that:

$$\begin{aligned}
Y_{10} &\sim Gamma(\frac{2\kappa_{11}\alpha_1}{\sigma_{11}^2}, \frac{\sigma_{11}^2}{2\kappa_{11}}) \text{ and} \\
Y_{20}|Y_{10} &\sim Gamma(\frac{2\kappa_{22}\{\alpha_2 + \frac{\kappa_{21}}{\kappa_{22}}(\alpha_1 - Y_{10})\}}{\sigma_{22}^2}, \frac{\sigma_{22}^2}{2\kappa_{22}}).
\end{aligned}$$

Let $v = \frac{4\kappa_{11}\alpha_1}{\sigma_{11}^2}, \lambda = cY_{1,t-1}e^{\kappa_{11}\triangle}$, and $c = \frac{4\kappa_{11}}{\sigma_{11}^2(1-e^{-\kappa_{11}\triangle})}$. Then, $cY_{1t}|Y_{1,t-1} \sim \chi_v^2(\lambda)$.
Write $w = \frac{4\kappa_{22}\{\alpha_2 + \frac{\kappa_{21}}{\kappa_{22}}(\alpha_1 - Y_{10})\}}{\sigma_{22}^2}, \eta = bY_{2,t-1}e^{\kappa_{22}\triangle}$, and $b = \frac{4\kappa_{22}}{\sigma_{22}^2(1-e^{-\kappa_{22}\triangle})}$. Then

$$bY_{2t}|Y_{t-1} \sim \chi_w^2(\eta).$$

If $Y \sim Gamma(k, \theta)$, then $Y^{-1} \sim$ Inverse-Gamma(k, θ^{-1}) with density

$$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}exp(\frac{-\beta}{x}), \;\; x > 0.$$

Following Theorem 2.2.3, we derive the estimated bias and variance for these interested parameters with the aid of Mathematica. The order of the two dimensional case turns out to be the same as the one dimensional case. Simulation results are provided in Section 2.3 to help appreciate the result.

2.3 Visualization of The Theoretical Biases and Variances

In this section, we plot the theoretical biases when related to the parameters for two dimensional O-U process driven by independent Brownian motions.

We can see from the following plots to get some intuitive feeling about how the biases are influenced by each other.

In Figure 2.1, the biases of $\hat{\kappa}_{11}$ and σ_{11}^2 are plotted separately. Note that the estimators for those two have the same properties as the one dimensional O-U process case. This is because in our specification, the first process in driven by only one independent Brownian motion, while the second process is driven by another independent Brownian motion and influenced by the first process. The main order of the bias for κ_{11} is $O(\frac{1}{n\triangle})$, and hence it is mainly influenced by sample size and sampling interval, and changes little with κ_{11}. The major order for $\hat{\sigma}_{11}^2$ is $O(\frac{1}{n})$, hence we get better estimation for σ_{11}^2 than the reversion parameter κ_{11}. The bias for $\hat{\sigma}_{11}^2$ is maily influenced by sample size, but also become larger when the true value σ_{11}^2 increases.

We plot the bias of $\hat{\kappa}_{22}$ in Figure 2.2 for three different values of κ_{21}, 0, 0.55, and 0.8. We did this to see intuitively how the correlation between the first process and the second process may influence the estimation of κ_{22}. When $\kappa_{21} = 0$, the bias for κ_{22} is enlarged relative to a pure independent O-U process because we are estimating one more parameter. When κ_{21} is larger, say 0.8 rather than 0.5, the bias is influenced more by the true value of κ_{22} than κ_{11}. The bias of κ_{22} is influenced more by σ_{22}^2 than σ_{11}^2.

Figure 2.3 shows the bias for $\hat{\kappa}_{21}$ seems to be influenced by κ_{11} and κ_{22} relatively similarly when κ_{21} equals 0.55 and 0.8. The increase of the value of σ_{22}^2 changes the bias of κ_{21} slowly, while the value of σ_{11}^2 changes the bias of κ_{21} very quickly at beginning, especially when σ_{22}^2 is small.

Figure 2.4 illustrates the bias of $\hat{\sigma}_{22}^2$, which is not only influenced by κ_{22} and σ_{22}^2, but also $n, \kappa_{11}, \kappa_{21}$ and κ_{22}. When $\kappa_{21} = 0.55$, the influence to the bias of $\hat{\sigma}_{22}^2$ from κ_{11} changes faster than that from κ_{22}.

Figure 2.1 Estimation Biases of κ_{11} and σ_{11}^2 under Two Dimensional O-U process Driven by Independent Brownian Motions.

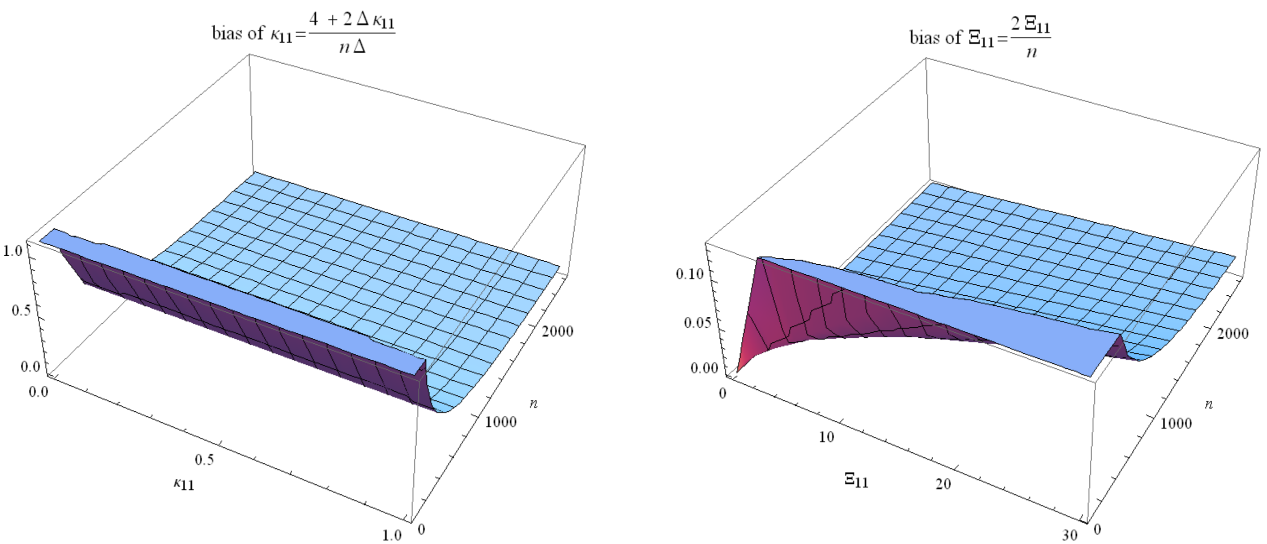


Figure 2.2 Estimation Bias of κ_{22} under Two Dimensional O-U process Driven by Independent Brownian Motions.

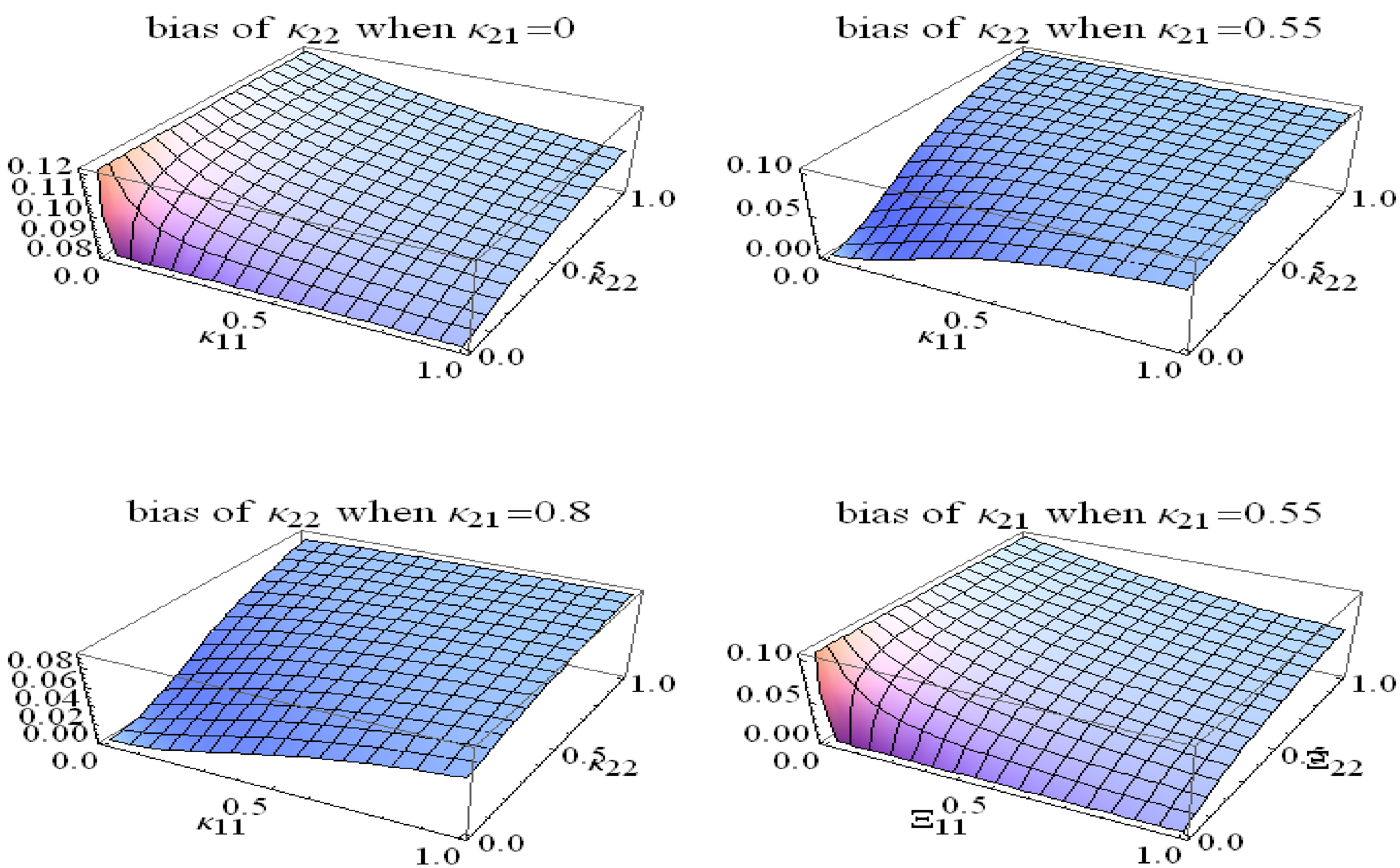


Figure 2.3 Estimation Bias of κ_{21} under Two Dimensional O-U process Driven by Independent Brownian Motions.

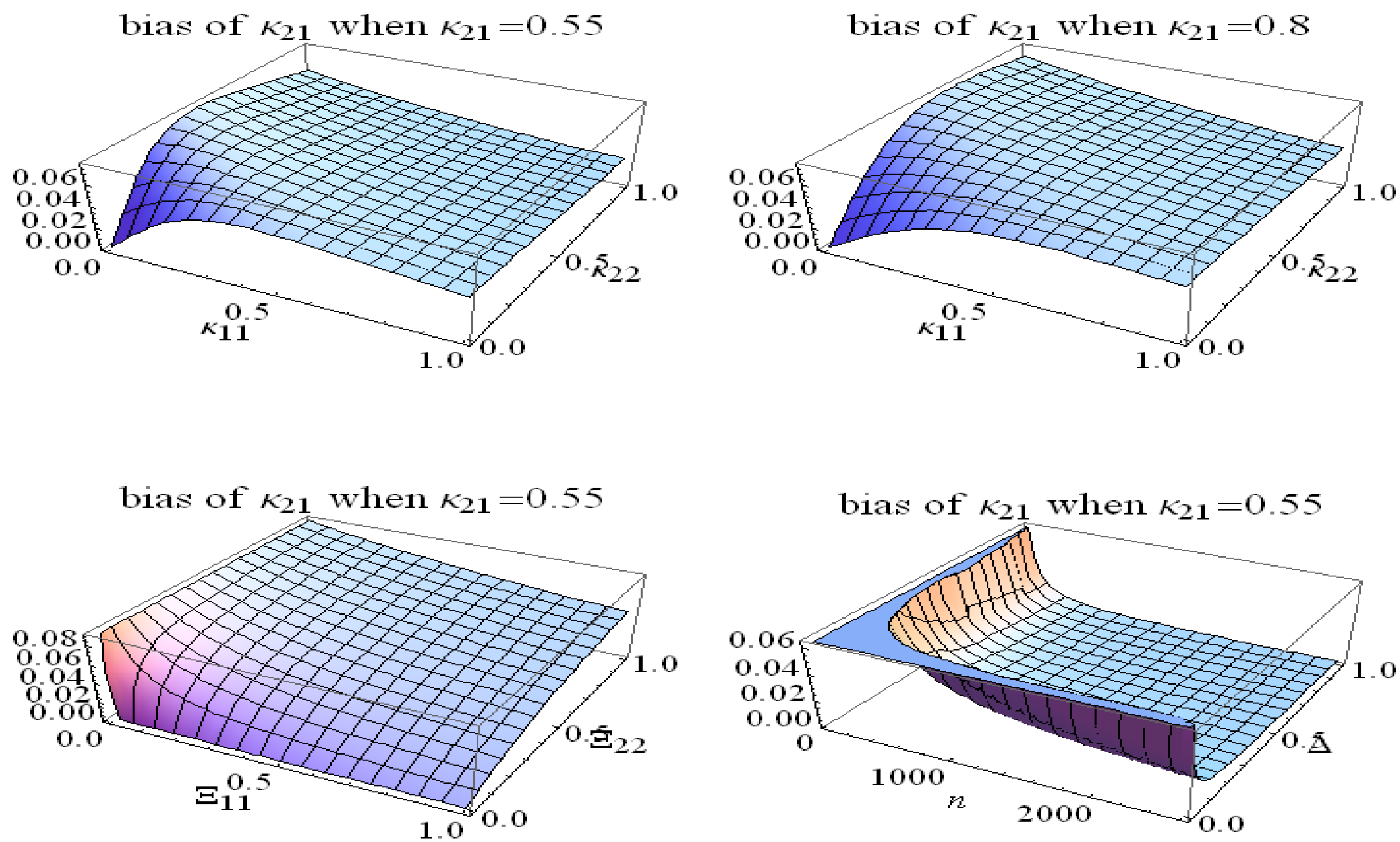
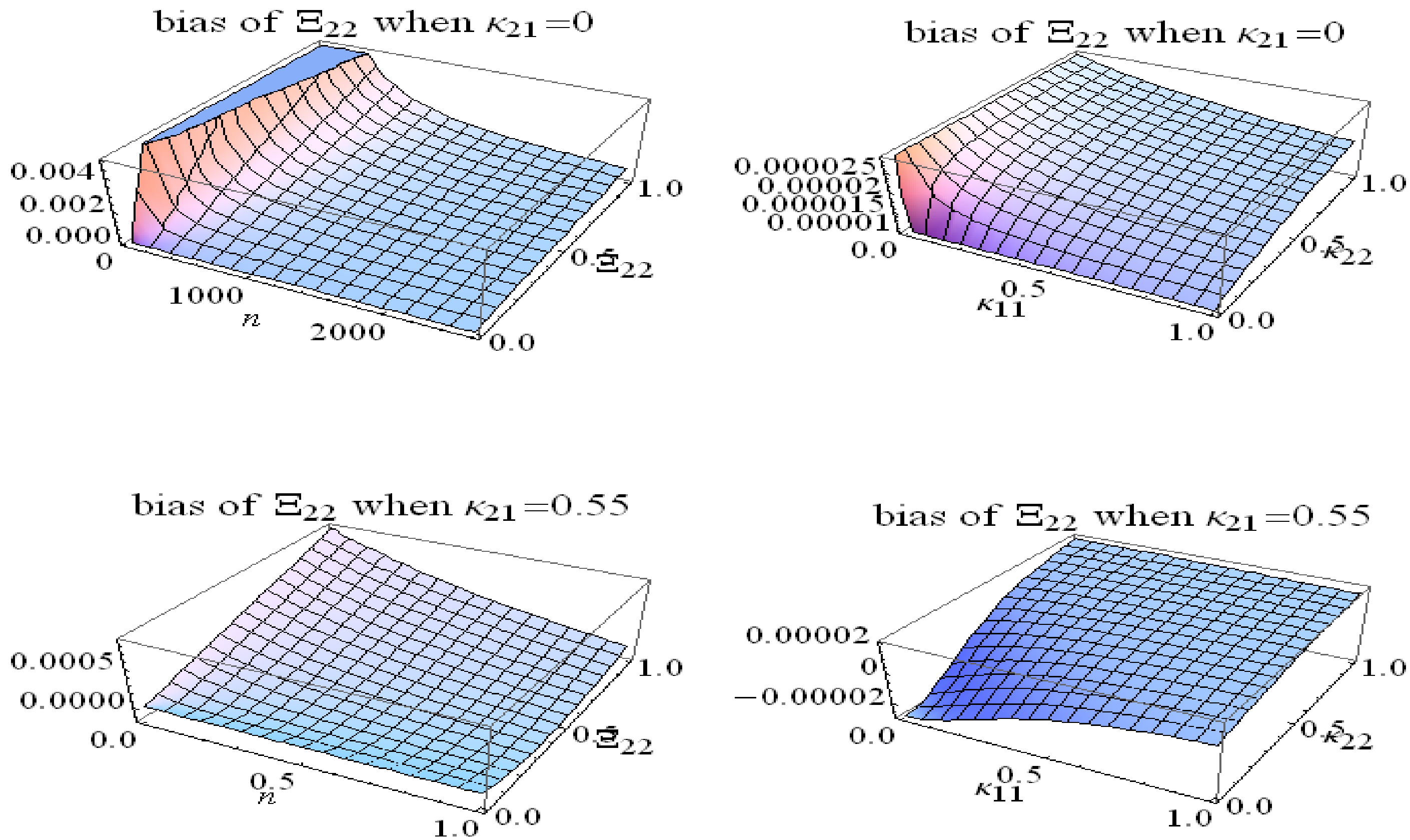


Figure 2.4 Estimation Bias of σ_{22}^2 under Two Dimensional O-U process Driven by Independent Brownian Motions..



2.4 Simulation Results

We consider the two dimensional Ornstein-Uhlenbeck diffusion process driven by independent Brownian motion specified in (1.4). The true parameter used for the following simulations are

$$\alpha_1 = 0.1, \alpha_2 = 0.12, \kappa_{11} = 0.25, \kappa_{22} = 0.4, \kappa_{12} = 0.55, \sigma_{11} = 0.015, \sigma_{22} = 0.025.$$

The simulation results are summarized in Table 2.1 to 2.10. For each parameter, we report the theoretical bias calculated from our theoretical result, the estimated value, the estimation bias, the relative bias in percentage, the mean square error, the estimated covariance matrix, as well as the bootstrapped estimation.

Supposed q years of sample is collected. When monthly sample is adopted, $\triangle = 1/12$ and the total sample size would be $12q$. When weekly sample is used, $\triangle = 1/52$ and the total sample size will become $52q$.

From the simulation results, we can see the estimation for the inversion parameter is much worse than the long term mean parameter and the variance parameter. Take the smallest sample size 10 year monthly data for example. The relative bias for the long term mean α_1 and α_2 are 4.70% and 16.58%, and are as small as 1.39% and 0.41% respectively for σ_{11}^2 and σ_{22}^2. But we see the relatively bias for κ_{11}, κ_{22} and κ_{12} are 165.36%, 122.15% and 51.00% respectively. We also see the relatively biases decrease as the sample sizes are increased to 20 year, 50 year, 100 year and 200 year. Note when 200 year monthly sample is adapted, the relatively biases for α_1, α_2, σ_{11}^2 and σ_{22}^2 are around 0.1%, and the biases for κ_{11}, κ_{22} and κ_{12} decreases to 8.14%, 4.89%, and 2.66%.

Similar simulation results are harvested for yearly data. The sample sizes used include 2 year, 6 year, 10 year, 20 year, and 200 year.

Six sets of simulation are provided:

Two Dimensional O-U Process Driven by Independent Brownian Motion with 5000 Simulations: Monthly Data;

Table 2.1 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 10 Year Monthly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.008333	0.008333	0.404167	0.370321	0.317333	0.000250	0.000137
Estimated Value	0.104699	0.139897	0.663411	0.888581	0.830496	0.015209	0.025104
Estimation Bias	0.004699	0.019897	0.413411	0.488581	0.280496	0.000209	0.000104
Relative Bias (%)	4.699310	16.580563	165.364319	122.145143	50.999323	1.392675	0.417015
Mean Square Error	0.019530	0.023775	0.423515	0.522201	0.468162	0.000004	0.000011
Estimated Variance	0.019507	-0.006513	0.004109	-0.003251	-0.004238	-0.000003	0.000004
	-0.006513	0.023379	-0.000724	-0.000411	-0.005615	0.000001	-0.000003
	0.004109	-0.000724	0.252606	-0.022530	-0.017593	0.000215	0.000010
	-0.003251	-0.000411	-0.022530	0.283490	0.088224	-0.000042	0.000388
	-0.004238	-0.005615	-0.017593	0.088224	0.389484	-0.000056	0.000047
	-0.000003	0.000001	0.000215	-0.000042	-0.000056	0.000004	0.000000
	0.000004	-0.000003	0.000010	0.000388	0.000047	0.000000	0.000011

Table 2.2 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 20 Year Monthly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.004167	0.004167	0.202083	0.185161	0.158666	0.000125	0.000069
Estimated Value	0.098991	0.130855	0.460532	0.622781	0.695077	0.015105	0.025036
Estimation Bias	-0.001009	0.010855	0.210532	0.222781	0.145077	0.000105	0.000036
Relative Bias (%)	-1.009010	9.046117	84.212949	55.695142	26.377691	0.702988	0.144703
Mean Square Error	0.006190	0.014558	0.116809	0.129411	0.153027	0.000002	0.000006
Estimated Variance	0.006189	-0.006123	0.002138	-0.001274	0.000160	2.0e-06	0.000002
	-0.006123	0.014440	-0.000880	0.000139	-0.001596	-3.0e-06	-0.000005
	0.002138	-0.000880	0.072485	-0.002901	-0.006125	6.7e-05	-0.000008
	-0.001274	0.000139	-0.002901	0.079780	0.032005	-5.0e-06	0.000134
	0.000160	-0.001596	-0.006125	0.032005	0.131980	-2.5e-05	0.000040
	0.000002	-0.000003	0.000067	-0.000005	-0.000025	2.0e-06	0.000000
	0.000002	-0.000005	-0.000008	0.000134	0.000040	0.0e+00	0.000006

Table 2.3 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 50 Year Monthly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.001667	0.001667	0.080833	0.074064	0.063467	0.000050	0.000027
Estimated Value	0.097970	0.124981	0.334144	0.482397	0.608733	0.015045	0.024992
Estimation Bias	-0.002030	0.004981	0.084144	0.082397	0.058733	0.000045	-0.000008
Relative Bias (%)	-2.030420	4.150968	33.657538	20.599141	10.678776	0.298300	-0.032255
Mean Square Error	0.003491	0.008382	0.024606	0.024941	0.039072	0.000001	0.000002
Estimated Variance	0.003487	-0.004318	0.000188	-0.000017	0.000305	1.0e-06	-1.0e-06
	-0.004318	0.008357	-0.000066	-0.000011	-0.000454	-1.0e-06	1.0e-06
	0.000188	-0.000066	0.017526	0.000065	-0.000467	1.5e-05	-2.0e-06
	-0.000017	-0.000011	0.000065	0.018151	0.010825	-2.0e-06	3.0e-05
	0.000305	-0.000454	-0.000467	0.010825	0.035622	-3.0e-06	1.7e-05
	0.000001	-0.000001	0.000015	-0.000002	-0.000003	1.0e-06	0.0e+00
	-0.000001	0.000001	-0.000002	0.000030	0.000017	0.0e+00	2.0e-06

Table 2.4 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 100 Year Monthly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.000833	0.000833	0.040417	0.037032	0.031733	0.000025	0.000014
Estimated Value	0.099224	0.121218	0.290981	0.439566	0.579390	0.015020	0.024998
Estimation Bias	-0.000776	0.001218	0.040981	0.039566	0.029390	0.000020	-0.000002
Relative Bias (%)	-0.776179	1.014947	16.392364	9.891590	5.343581	0.132839	-0.006974
Mean Square Error	0.002137	0.005219	0.008587	0.008784	0.016042	0.000000	0.000001
Estimated Variance	0.002136	-0.002809	0.000025	-0.000020	0.000041	0e+00	0.0e+00
	-0.002809	0.005218	-0.000018	0.000095	0.000019	0e+00	0.0e+00
	0.000025	-0.000018	0.006907	0.000016	-0.000195	7e-06	0.0e+00
	-0.000020	0.000095	0.000016	0.007218	0.005113	0e+00	1.4e-05
	0.000041	0.000019	-0.000195	0.005113	0.015178	0e+00	1.2e-05
	0.000000	0.000000	0.000007	0.000000	0.000000	0e+00	0.0e+00
	0.000000	0.000000	0.000000	0.000014	0.000012	0e+00	1.0e-06

Table 2.5 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 200 Year Monthly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.000417	0.000417	0.020208	0.018516	0.015867	0.000013	6.864e-6
Estimated Value	0.100078	0.120075	0.270354	0.419571	0.564638	0.015019	0.024995
Estimation Bias	0.000078	0.000075	0.020354	0.019571	0.014638	0.000019	-0.000005
Relative Bias (%)	0.078054	0.062505	8.141537	4.892728	2.661393	0.125813	-0.021901
Mean Square Error	0.001172	0.002953	0.003350	0.003729	0.007311	0.000000	0.000001
Estimated Variance	0.001172	-0.001592	-0.000003	-0.000004	0.000084	0e+00	0e+00
	-0.001592	0.002953	0.000018	0.000021	-0.000095	0e+00	1e-06
	-0.000003	0.000018	0.002936	0.000030	-0.000044	3e-06	1e-06
	-0.000004	0.000021	0.000030	0.003346	0.002704	0e+00	6e-06
	0.000084	-0.000095	-0.000044	0.002704	0.007097	0e+00	6e-06
	0.000000	0.000000	0.000003	0.000000	0.000000	0e+00	0e+00
	0.000000	0.000001	0.000001	0.000006	0.000006	0e+00	1e-06

Table 2.6 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 2 Year Weekly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.009615	0.009615	2.004808	1.834579	1.556933	0.000288	0.000158
Estimated Value	0.144112	0.176726	2.084614	2.963584	1.882013	0.015105	0.025131
Estimation Bias	0.044112	0.056726	1.834614	2.563584	1.332013	0.000105	0.000131
Relative Bias (%)	44.112467	47.271635	733.845698	640.896086	242.184119	0.697598	0.522589
Mean Square Error	0.102186	0.112417	8.558545	13.170693	7.002077	0.000004	0.000013
Estimated Variance	0.100240	0.002005	-0.040537	0.068447	-0.056085	-0.000012	0.000018
	0.002005	0.109199	0.025768	-0.015135	-0.039392	0.000008	-0.000021
	-0.040537	0.025768	5.192735	-0.165321	0.317856	0.000762	0.000102
	0.068447	-0.015135	-0.165321	6.598729	1.681385	-0.000145	0.001902
	-0.056085	-0.039392	0.317856	1.681385	5.227819	-0.000165	0.000449
	-0.000012	0.000008	0.000762	-0.000145	-0.000165	0.000004	0.000000
	0.000018	-0.000021	0.000102	0.001902	0.000449	0.000000	0.000013
BT Estimation	0.120359	0.147340	0.535871	1.023182	0.769530	0.014949	0.024953
BT Estimation Bias	0.020359	0.027340	0.285871	0.623182	0.219530	-0.000051	-0.000047
BT Relative Bias (%)	20.35913	22.78321	114.34841	155.79562	39.91446	-0.34273	-0.18799
BT Mean Square Error	0.111020	0.124049	4.454205	6.428589	5.675905	0.000004	0.000013

Table 2.7 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 6 Year Weekly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.003205	0.003205	0.668269	0.611526	0.518978	0.000096	0.000053
Estimated Value	0.112586	0.154396	0.880956	1.222616	0.970725	0.015038	0.025043
Estimation Bias	0.012586	0.034396	0.630956	0.822616	0.420725	0.000038	0.000043
Relative Bias (%)	12.586044	28.663629	252.382512	205.654112	76.495441	0.252092	0.173069
Mean Square Error	0.022238	0.070817	0.954082	1.381254	0.947118	0.000001	0.000004
Estimated Variance	0.022080	-0.004450	-0.001009	0.009332	-0.004140	-5.0e-06	-0.000002
	-0.004450	0.069634	-0.000389	-0.007717	-0.016972	-5.0e-06	-0.000009
	-0.001009	-0.000389	0.555977	-0.022970	0.018480	7.7e-05	-0.000019
	0.009332	-0.007717	-0.022970	0.704556	0.161098	-1.4e-05	0.000219
	-0.004140	-0.016972	0.018480	0.161098	0.770109	-2.5e-05	0.000052
	-0.000005	-0.000005	0.000077	-0.000014	-0.000025	1.0e-06	0.000000
	-0.000002	-0.000009	-0.000019	0.000219	0.000052	0.0e+00	0.000004
BT Estimation	0.100134	0.132644	0.333596	0.565224	0.595372	0.014977	0.024977
BT Estimation Bias	0.000134	0.012644	0.083596	0.165224	0.045372	-0.000023	-0.000023
BT Relative Bias	0.134011	10.536686	33.438443	41.305885	8.249421	-0.151871	-0.091808
BT Mean Square Error	0.028156	0.080379	0.484919	0.673025	0.804052	0.000001	0.000004

Table 2.8 Estimation Sesults for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 10 Year Weekly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.001923	0.001923	0.400962	0.366916	0.311387	0.000058	0.000032
Estimated Value	0.102946	0.138688	0.632105	0.847540	0.802252	0.015037	0.025006
Estimation Bias	0.002946	0.018688	0.382105	0.447540	0.252252	0.000037	0.000006
Relative Bias (%)	2.945964	15.573231	152.842031	111.884995	45.864031	0.248194	0.022848
Mean Square Error	0.010234	0.023073	0.364348	0.456574	0.395509	0.000001	0.000002
Estimated Variance	0.010225	-0.005697	0.002379	0.001335	-0.000138	-1.0e-06	-0.000001
	-0.005697	0.022724	0.001716	0.000222	-0.007028	2.0e-06	-0.000007
	0.002379	0.001716	0.218344	-0.001318	-0.003968	4.1e-05	0.000000
	0.001335	0.000222	-0.001318	0.256282	0.063991	-4.0e-06	0.000106
	-0.000138	-0.007028	-0.003968	0.063991	0.331877	1.0e-06	0.000021
	-0.000001	0.000002	0.000041	-0.000004	0.000001	1.0e-06	0.000000
	-0.000001	-0.000007	0.000000	0.000106	0.000021	0.0e+00	0.000002
BT Estimation	0.095370	0.120260	0.297201	0.457274	0.572205	0.014999	0.024968
BT Estimation Bias	-0.004630	0.000260	0.047201	0.057274	0.022205	-0.000001	-0.000032
BT Relative Bias	-4.630380	0.216766	18.880480	14.318596	4.037346	-0.009763	-0.129196
BT Mean Square Error	0.014531	0.032232	0.193921	0.238235	0.339362	0.000001	0.000002

Table 2.9 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 20 Year Weekly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.000962	0.000962	0.200481	0.183458	0.155693	0.000029	0.000016
Estimated Value	0.098885	0.129708	0.453556	0.612892	0.682946	0.015024	0.025010
Estimation Bias	-0.001115	0.009708	0.203556	0.212892	0.132946	0.000024	0.000010
Relative Bias (%)	-1.115103	8.089987	81.422452	53.223060	24.171912	0.158146	0.038718
Mean Square Error	0.005954	0.013712	0.109564	0.120168	0.139218	0.000000	0.000001
Estimated Variance	0.005953	-0.005783	0.001120	-0.000758	0.000157	-1.0e-06	0.0e+00
	-0.005783	0.013618	0.000211	0.000337	-0.002653	0.0e+00	0.0e+00
	0.001120	0.000211	0.068129	0.000023	-0.005517	1.1e-05	-4.0e-06
	-0.000758	0.000337	0.000023	0.074845	0.027860	-2.0e-06	2.6e-05
	0.000157	-0.002653	-0.005517	0.027860	0.121543	3.0e-06	1.0e-05
	-0.000001	0.000000	0.000011	-0.000002	0.000003	0.0e+00	0.0e+00
	0.000000	0.000000	-0.000004	0.000026	0.000010	0.0e+00	1.0e-06

Table 2.10 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Independent Brownian Motion with 200 Year Weekly Sample, 5000 Simulations.

Parameters	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025
Theoretical Bias	0.000096	0.000096	0.020048	0.018346	0.015569	2.885e-6	1.584e-6
Estimated Value	0.099589	0.120550	0.268860	0.418144	0.563913	0.015007	0.025006
Estimation Bias	-0.000411	0.000550	0.018860	0.018144	0.013913	0.000007	0.000006
Relative Bias (%)	-0.410505	0.458492	7.544129	4.536065	2.529633	0.047518	0.022280
Mean Square Error	0.001153	0.002891	0.003063	0.003462	0.006519	0.000000	0.000000
Estimated Variance	0.001153	-0.001564	0.000019	0.000035	-0.000001	0e+00	0e+00
	-0.001564	0.002891	-0.000002	-0.000049	-0.000062	0e+00	0e+00
	0.000019	-0.000002	0.002708	0.000049	-0.000137	1e-06	0e+00
	0.000035	-0.000049	0.000049	0.003133	0.002256	0e+00	1e-06
	-0.000001	-0.000062	-0.000137	0.002256	0.006325	0e+00	2e-06
	0.000000	0.000000	0.000001	0.000000	0.000000	0e+00	0e+00
	0.000000	0.000000	0.000000	0.000001	0.000002	0e+00	0e+00

Two Dimensional O-U Process Driven by Independent Brownian Motion with 5000 Simulations: Weekly Data.

Two Dimensional O-U Process Driven by Correlated Brownian Motion with 5000 Simulations: Monthly Data;

Two Dimensional O-U Process Driven by Correlated Brownian Motion with 5000 Simulations: Weekly Data;

Two Dimensional CIR Process Driven by Independent Brownian Motion with 5000 Simulations: Monthly Data;

Two Dimensional CIR Process Driven by Independent Brownian Motion with 5000 Simulations: Weekly Data.

Comparing those estimation results for monthly data and weekly data with the same sampling years, the variation of the data is relatively small from week to week and hence may be unable to provide enough information to enable good estimation even with relatively large sample size, say 4 years. For future estimation purposes, we better collect monthly data than weekly data.

We also provide the simulation results for the two dimensional Ornstein-Uhlenbeck diffusion process. The parameters adopted are the same as in the independent case, just with one more correlation parameter ρ, which is chosen to be 0.25. We still see worse estimation for reversion parameters than the others, but the estimations for κ_{22} do seem to be improved a little. We also observe that for the two dimensional O-U process driven by correlated Brownian motions, the estimation effect for κ_{22} is generally improved than the one driven by independent Brownian motion. However, the estimation for volatility σ_{22} is not as good as the independent version when weekly data with small sample is used. The reason can be traced back to the original form of the correlated process. Succinctly, when we convert the correlated form to the equivalent independent form, σ_{22} and ρ co-exist in the second diffusion term for the second process as $\sigma_{22}\sqrt{1-\rho^2}$. Hence, the difficulty for estimating σ_{22} is increased.

As for the two dimensional CIR process, under the same specification for parameters and

Table 2.11 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 10 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025	0.250
Estimated Value	0.106642	0.131605	0.509657	0.452988	0.615382	0.015055	0.024747	0.247622
Estimation Bias	0.006642	0.011605	0.259657	0.052988	0.065382	0.000055	-0.000253	-0.002378
Relative Bias (%)	6.641733	9.670643	103.862936	13.246937	11.887597	0.365956	-1.012571	-0.951342
Mean Square Error	0.013993	0.013207	0.264257	0.092776	0.150653	0.000004	0.000010	0.005372
Estimated Variance	0.013949	-0.004242	0.000645	0.003400	-0.001088	-0.000005	7.0e-06	-0.000519
	-0.004242	0.013073	0.000670	0.003033	-0.000944	-0.000007	5.0e-06	0.000142
	0.000645	0.000670	0.196835	0.043284	0.054024	0.000113	8.8e-05	-0.004949
	0.003400	0.003033	0.043284	0.089968	-0.004907	-0.000013	1.4e-04	-0.005576
	-0.001088	-0.000944	0.054024	-0.004907	0.146378	0.000008	3.0e-06	-0.000893
	-0.000005	-0.000007	0.000113	-0.000013	0.000008	0.000004	0.0e+00	0.000017
	0.000007	0.000005	0.000088	0.000140	0.000003	0.000000	1.0e-05	0.000038
	-0.000519	0.000142	-0.004949	-0.005576	-0.000893	0.000017	3.8e-05	0.005366

Table 2.12 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 20 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025	0.25
Estimated Value	0.100104	0.128044	0.436926	0.439098	0.612040	0.015064	0.024866	0.250207
Estimation Bias	0.000104	0.008044	0.186926	0.039098	0.062040	0.000064	-0.000134	0.000207
Relative Bias (%)	0.103796	6.703699	74.770374	9.774418	11.280043	0.426316	-0.534818	0.082908
Mean Square Error	0.006086	0.011447	0.105208	0.030094	0.083534	0.000002	0.000005	0.003627
Estimated Variance	0.006086	-0.005346	0.000946	-0.000053	0.000453	3.0e-06	0.000000	0.000109
	-0.005346	0.011382	-0.000031	0.000800	-0.000829	-3.0e-06	0.000003	-0.000187
	0.000946	-0.000031	0.070267	0.005212	0.020917	6.3e-05	0.000015	-0.000668
	-0.000053	0.000800	0.005212	0.028565	0.001014	2.7e-05	0.000107	-0.001233
	0.000453	-0.000829	0.020917	0.001014	0.079685	3.3e-05	0.000045	0.001512
	0.000003	-0.000003	0.000063	0.000027	0.000033	2.0e-06	0.000000	0.000015
	0.000000	0.000003	0.000015	0.000107	0.000045	0.0e+00	0.000005	0.000026
	0.000109	-0.000187	-0.000668	-0.001233	0.001512	1.5e-05	0.000026	0.003627

Table 2.13 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 50 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025	0.25
Estimated Value	0.101641	0.120304	0.314568	0.364048	0.543895	0.015048	0.024893	0.252393
Estimation Bias	0.001641	0.000304	0.064568	-0.035952	-0.006105	0.000048	-0.000107	0.002393
Relative Bias (%)	1.641398	0.253351	25.827133	-8.987920	-1.110009	0.319196	-0.429316	0.957378
Mean Square Error	0.003276	0.005692	0.018972	0.005142	0.017947	0.000001	0.000002	0.001417
Estimated Variance	0.003273	-0.003315	0.000044	-0.000044	0.000136	1e-06	1.0e-06	-0.000017
	-0.003315	0.005692	0.000024	0.000191	-0.000061	0e+00	-2.0e-06	0.000031
	0.000044	0.000024	0.014803	0.000521	0.005213	8e-06	1.0e-06	-0.000469
	-0.000044	0.000191	0.000521	0.003849	0.001779	-5e-06	1.0e-05	-0.000552
	0.000136	-0.000061	0.005213	0.001779	0.017910	-1e-06	1.1e-05	0.000015
	0.000001	0.000000	0.000008	-0.000005	-0.000001	1e-06	0.0e+00	0.000006
	0.000001	-0.000002	0.000001	0.000010	0.000011	0e+00	2.0e-06	0.000010
	-0.000017	0.000031	-0.000469	-0.000552	0.000015	6e-06	1.0e-05	0.001411

Table 2.14 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 100 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025	0.25
Estimated Value	0.100495	0.120325	0.288909	0.388425	0.557423	0.015030	0.024951	0.249614
Estimation Bias	0.000495	0.000325	0.038909	-0.011575	0.007423	0.000030	-0.000049	-0.000386
Relative Bias (%)	0.494731	0.271029	15.563744	-2.893749	1.349657	0.202616	-0.195870	-0.154370
Mean Square Error	0.002157	0.004235	0.008097	0.004911	0.011637	0.000000	0.000001	0.000749
Estimated Variance	0.002157	-0.002437	-0.000023	0.000117	-0.000046	0.0e+00	0.0e+00	-0.000005
	-0.002437	0.004235	0.000070	-0.000112	0.000205	0.0e+00	1.0e-06	0.000018
	-0.000023	0.000070	0.006583	0.000120	0.002285	7.0e-06	3.0e-06	0.000015
	0.000117	-0.000112	0.000120	0.004777	0.002665	1.4e-05	4.2e-05	-0.000228
	-0.000046	0.000205	0.002285	0.002665	0.011581	1.0e-05	3.3e-05	0.000108
	0.000000	0.000000	0.000007	0.000014	0.000010	0.0e+00	0.0e+00	0.000003
	0.000000	0.000001	0.000003	0.000042	0.000033	0.0e+00	1.0e-06	0.000006
	-0.000005	0.000018	0.000015	-0.000228	0.000108	3.0e-06	6.0e-06	0.000748

Table 2.15 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 200 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.1	0.12	0.25	0.4	0.55	0.015	0.025	0.25
Estimated Value	0.099945	0.120346	0.268205	0.341370	0.516265	0.014992	0.024836	0.241679
Estimation Bias	-0.000055	0.000346	0.018205	-0.058630	-0.033735	-0.000008	-0.000164	-0.008321
Relative Bias (%)	-0.054583	0.287955	7.282091	-14.657615	-6.133567	-0.054652	-0.654227	-3.328395
Mean Square Error	0.001229	0.002249	0.003051	0.003776	0.004758	0.000000	0.000001	0.000882
Estimated Variance	0.001229	-0.001285	-0.000019	-0.000008	-0.000101	0e+00	-1e-06	0.000006
	-0.001285	0.002249	-0.000004	0.000006	0.000119	0e+00	1e-06	-0.000003
	-0.000019	-0.000004	0.002720	0.000017	0.001000	2e-06	-1e-06	-0.000274
	-0.000008	0.000006	0.000017	0.000339	0.000224	0e+00	2e-06	-0.000004
	-0.000101	0.000119	0.001000	0.000224	0.003620	1e-06	3e-06	-0.000002
	0.000000	0.000000	0.000002	0.000000	0.000001	0e+00	0e+00	0.000003
	-0.000001	0.000001	-0.000001	0.000002	0.000003	0e+00	1e-06	0.000005
	0.000006	-0.000003	-0.000274	-0.000004	-0.000002	3e-06	5e-06	0.000813

Table 2.16 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 2 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025	0.250
Estimated Value	0.115171	0.209770	1.032175	0.384038	0.814093	0.017934	0.027295	0.264424
Estimation Bias	0.015171	0.089770	0.782175	-0.015962	0.264093	0.002934	0.002295	0.014424
Relative Bias (%)	15.170979	74.808644	312.869978	-3.990418	48.016862	19.558424	9.178335	5.769785
Mean Square Error	0.016649	0.127573	3.120618	0.381799	1.052504	0.000054	0.000045	0.006692
Estimated Variance	0.016419	0.000660	0.017343	0.004632	-0.001852	-0.000051	-0.000028	-0.000269
	0.000660	0.119515	0.056616	-0.007563	0.070800	-0.000259	-0.000225	-0.001811
	0.017343	0.056616	2.508821	0.269368	0.672571	-0.001959	-0.002201	-0.017636
	0.004632	-0.007563	0.269368	0.381545	-0.081934	-0.000368	-0.000392	-0.007462
	-0.001852	0.070800	0.672571	-0.081934	0.982759	-0.000898	-0.000983	-0.003539
	-0.000051	-0.000259	-0.001959	-0.000368	-0.000898	0.000045	0.000034	0.000133
	-0.000028	-0.000225	-0.002201	-0.000392	-0.000983	0.000034	0.000040	0.000148
	-0.000269	-0.001811	-0.017636	-0.007462	-0.003539	0.000133	0.000148	0.006484

Table 2.17 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 6 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025	0.250
Estimated Value	0.114045	0.195141	0.732099	0.236298	0.684078	0.015128	0.024981	0.254038
Estimation Bias	0.014045	0.075141	0.482099	-0.163702	0.134078	0.000128	-0.000019	0.004038
Relative Bias (%)	14.044612	62.617881	192.839633	-40.925561	24.377746	0.853885	-0.075965	1.615062
Mean Square Error	0.014003	0.118732	0.726641	0.082715	0.318812	0.000002	0.000007	0.002858
Estimated Variance	0.013806	-0.005059	0.002978	0.000939	-0.003073	-4.0e-06	4.0e-06	-0.000136
	-0.005059	0.113085	0.017927	-0.007245	0.013281	-4.0e-06	4.1e-05	-0.000024
	0.002978	0.017927	0.494221	0.019364	0.150982	1.2e-05	-4.9e-05	-0.001852
	0.000939	-0.007245	0.019364	0.055917	-0.017629	-4.0e-06	-3.3e-05	-0.001250
	-0.003073	0.013281	0.150982	-0.017629	0.300835	1.0e-06	-1.9e-05	0.000843
	-0.000004	-0.000004	0.000012	-0.000004	0.000001	2.0e-06	1.0e-06	0.000015
	0.000004	0.000041	-0.000049	-0.000033	-0.000019	1.0e-06	7.0e-06	0.000031
	-0.000136	-0.000024	-0.001852	-0.001250	0.000843	1.5e-05	3.1e-05	0.002842

Table 2.18 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 10 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025	0.250
Estimated Value	0.103833	0.136358	0.510504	0.269028	0.611702	0.018229	0.027510	0.263248
Estimation Bias	0.003833	0.016358	0.260504	-0.130972	0.061702	0.003229	0.002510	0.013248
Relative Bias (%)	3.833209	13.631297	104.201536	-32.742943	11.218570	21.526980	10.038937	5.299122
Mean Square Error	0.006588	0.016209	0.228380	0.026112	0.140032	0.000055	0.000038	0.002038
Estimated Variance	0.006573	-0.003860	0.002569	0.000116	-0.000134	-0.000014	-0.000011	-0.000012
	-0.003860	0.015941	0.001811	-0.000727	0.001384	-0.000051	-0.000046	-0.000294
	0.002569	0.001811	0.160518	0.001795	0.061359	-0.000800	-0.000794	-0.002503
	0.000116	-0.000727	0.001795	0.008958	-0.002537	-0.000041	-0.000054	-0.000227
	-0.000134	0.001384	0.061359	-0.002537	0.136225	-0.000348	-0.000336	-0.000188
	-0.000014	-0.000051	-0.000800	-0.000041	-0.000348	0.000044	0.000035	0.000127
	-0.000011	-0.000046	-0.000794	-0.000054	-0.000336	0.000035	0.000032	0.000120
	-0.000012	-0.000294	-0.002503	-0.000227	-0.000188	0.000127	0.000120	0.001862

Table 2.19 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 20 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025	0.250
Estimated Value	0.100009	0.120331	0.250665	0.048508	0.500158	0.014396	0.024482	0.300930
Estimation Bias	0.000009	0.000331	0.000665	-0.351492	-0.049842	-0.000604	-0.000518	0.050930
Relative Bias (%)	0.009422	0.275712	0.266127	-87.873051	-9.062159	-4.026125	-2.073672	20.371960
Mean Square Error	0.000032	0.000137	0.000321	0.124176	0.002595	0.000001	0.000002	0.002597
Estimated Variance	3.2e-05	-0.000013	-0.000001	1.0e-06	-0.000012	0e+00	0.0e+00	1.0e-06
	-1.3e-05	0.000137	0.000046	3.2e-05	0.000016	0e+00	1.0e-06	2.0e-06
	-1.0e-06	0.000046	0.000321	6.2e-05	0.000070	0e+00	1.0e-06	6.0e-06
	1.0e-06	0.000032	0.000062	6.3e-04	0.000012	2e-06	2.4e-05	1.3e-05
	-1.2e-05	0.000016	0.000070	1.2e-05	0.000111	0e+00	0.0e+00	0.0e+00
	0.0e+00	0.000000	0.000000	2.0e-06	0.000000	0e+00	0.0e+00	0.0e+00
	0.0e+00	0.000001	0.000001	2.4e-05	0.000000	0e+00	2.0e-06	0.0e+00
	1.0e-06	0.000002	0.000006	1.3e-05	0.000000	0e+00	0.0e+00	3.0e-06

Table 2.20 Estimation Results for Two Dimensional O-U Process Driven by Two Dimensional Correlated Brownian Motion with 200 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}	ρ
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025	0.250
Estimated Value	0.099098	0.148072	0.265362	0.192709	0.432810	0.015313	0.024886	0.244539
Estimation Bias	-0.000902	0.028072	0.015362	-0.207291	-0.117190	0.000313	-0.000114	-0.005461
Relative Bias (%)	-0.902187	23.393225	6.144840	-51.822805	-21.307241	2.089704	-0.456912	-2.184445
Mean Square Error	0.001280	0.042672	0.003382	0.051054	0.020628	0.000005	0.000003	0.001066
Estimated Variance	0.001280	-0.001398	0.000035	-0.000007	-0.000026	0.0e+00	0.0e+00	0.000008
	-0.001398	0.041884	-0.000347	-0.003717	-0.002470	-9.0e-06	5.0e-06	0.000075
	0.000035	-0.000347	0.003146	-0.000052	0.001368	-5.0e-06	8.0e-06	0.000295
	-0.000007	-0.003717	-0.000052	0.008084	0.003486	1.8e-05	1.0e-05	-0.000169
	-0.000026	-0.002470	0.001368	0.003486	0.006895	2.0e-05	0.0e+00	0.000028
	0.000000	-0.000009	-0.000005	0.000018	0.000020	5.0e-06	4.0e-06	0.000015
	0.000000	0.000005	0.000008	0.000010	0.000000	4.0e-06	3.0e-06	0.000025
	0.000008	0.000075	0.000295	-0.000169	0.000028	1.5e-05	2.5e-05	0.001036

Table 2.21 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 10 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.112856	0.131191	0.751291	0.991848	-0.513873	0.015266	0.025572
Estimation Bias	0.012856	0.011191	0.501291	0.591848	-0.263873	0.000266	0.000572
Relative Bias (%)	12.855708	9.325770	200.516410	147.962031	105.549197	1.770811	2.289112
Mean Square Error	0.015446	0.011502	0.503991	0.702179	0.424969	0.000004	0.000013
Estimated Variance	0.015281	0.006214	-0.012895	0.001826	0.001860	-0.000012	0.000004
	0.006214	0.011376	-0.006520	-0.010643	-0.001300	-0.000006	-0.000010
	-0.012895	-0.006520	0.252699	-0.016116	-0.017493	0.000208	-0.000022
	0.001826	-0.010643	-0.016116	0.351895	-0.076899	-0.000008	0.000527
	0.001860	-0.001300	-0.017493	-0.076899	0.355340	0.000007	-0.000095
	-0.000012	-0.000006	0.000208	-0.000008	0.000007	0.000004	0.000000
	0.000004	-0.000010	-0.000022	0.000527	-0.000095	0.000000	0.000012

Table 2.22 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 20 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.103120	0.122758	0.480880	0.677429	-0.370193	0.015165	0.025355
Estimation Bias	0.003120	0.002758	0.230880	0.277429	-0.120193	0.000165	0.000355
Relative Bias (%)	3.120112	2.298710	92.351958	69.357284	48.077386	1.097722	1.418765
Mean Square Error	0.001954	0.001997	0.125506	0.175015	0.125541	0.000002	0.000006
Estimated Variance	0.001944	0.001203	-0.003166	0.000012	0.000328	-2e-06	-0.000001
	0.001203	0.001989	-0.002093	-0.003671	-0.000524	-1e-06	-0.000004
	-0.003166	-0.002093	0.072200	-0.003028	-0.004479	6e-05	-0.000002
	0.000012	-0.003671	-0.003028	0.098048	-0.022998	3e-06	0.000166
	0.000328	-0.000524	-0.004479	-0.022998	0.111095	1e-06	-0.000031
	-0.000002	-0.000001	0.000060	0.000003	0.000001	2e-06	0.000000
	-0.000001	-0.000004	-0.000002	0.000166	-0.000031	0e+00	0.000006

Table 2.23 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 50 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.100618	0.120432	0.335128	0.501126	-0.286661	0.015114	0.025261
Estimation Bias	0.000618	0.000432	0.085128	0.101126	-0.036661	0.000114	0.000261
Relative Bias (%)	0.617752	0.360194	34.051108	25.281511	14.664380	0.761020	1.043010
Mean Square Error	0.000494	0.000563	0.024793	0.033157	0.031248	0.000001	0.000002
Estimated Variance	0.000494	0.000300	-0.001207	0.000003	0.000042	-1.0e-06	0.0e+00
	0.000300	0.000563	-0.000686	-0.001244	-0.000100	-1.0e-06	0.0e+00
	-0.001207	-0.000686	0.017547	-0.000310	-0.000537	1.8e-05	-2.0e-06
	0.000003	-0.001244	-0.000310	0.022931	-0.006347	3.0e-06	4.4e-05
	0.000042	-0.000100	-0.000537	-0.006347	0.029904	1.0e-06	-1.1e-05
	-0.000001	-0.000001	0.000018	0.000003	0.000001	1.0e-06	0.0e+00
	0.000000	0.000000	-0.000002	0.000044	-0.000011	0.0e+00	2.0e-06

Table 2.24 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 100 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.100305	0.120409	0.288987	0.449376	-0.266423	0.015080	0.025218
Estimation Bias	0.000305	0.000409	0.038987	0.049376	-0.016423	0.000080	0.000218
Relative Bias (%)	0.304980	0.341221	15.594835	12.344051	6.569020	0.534514	0.873935
Mean Square Error	0.000238	0.000281	0.008117	0.012356	0.012630	0.000000	0.000001
Estimated Variance	0.000237	0.000148	-0.000566	-0.000007	-0.000007	0e+00	0.0e+00
	0.000148	0.000281	-0.000344	-0.000666	0.000004	0e+00	0.0e+00
	-0.000566	-0.000344	0.006597	-0.000144	0.000161	8e-06	-1.0e-06
	-0.000007	-0.000666	-0.000144	0.009918	-0.003156	0e+00	2.1e-05
	-0.000007	0.000004	0.000161	-0.003156	0.012361	0e+00	-8.0e-06
	0.000000	0.000000	0.000008	0.000000	0.000000	0e+00	0.0e+00
	0.000000	0.000000	-0.000001	0.000021	-0.000008	0e+00	1.0e-06

Table 2.25 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 200 Year Monthly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.100354	0.120377	0.268223	0.424022	-0.256758	0.015070	0.025187
Estimation Bias	0.000354	0.000377	0.018223	0.024022	-0.006758	0.000070	0.000187
Relative Bias (%)	0.353601	0.314214	7.289083	6.005452	2.703058	0.468641	0.747001
Mean Square Error	0.000117	0.000141	0.003297	0.005039	0.005696	0.000000	0.000001
Estimated Variance	0.000117	0.000074	-0.000296	-0.000003	-0.000002	0e+00	0e+00
	0.000074	0.000141	-0.000172	-0.000320	0.000015	0e+00	0e+00
	-0.000296	-0.000172	0.002965	-0.000092	0.000065	3e-06	1e-06
	-0.000003	-0.000320	-0.000092	0.004462	-0.001614	0e+00	9e-06
	-0.000002	0.000015	0.000065	-0.001614	0.005651	0e+00	-3e-06
	0.000000	0.000000	0.000003	0.000000	0.000000	0e+00	0e+00
	0.000000	0.000000	0.000001	0.000009	-0.000003	0e+00	1e-06

Table 2.26 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 2 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.150230	0.167919	2.869705	3.541580	-1.755716	0.015216	0.025497
Estimation Bias	0.050230	0.047919	2.619705	3.141580	-1.505716	0.000216	0.000497
Relative Bias (%)	50.230060	39.932391	1047.881990	785.394905	602.286296	1.439197	1.986871
Mean Square Error	0.074713	0.068579	12.666537	17.271588	9.395479	0.000005	0.000014
Estimated Variance	0.072190	0.026841	-0.143674	0.083831	0.001107	-0.000014	0.000012
	0.026841	0.066283	-0.059748	-0.071326	-0.046554	-0.000009	-0.000016
	-0.143674	-0.059748	5.803683	-0.501743	-0.485704	0.001046	-0.000082
	0.083831	-0.071326	-0.501743	7.402065	-1.733532	-0.000211	0.001946
	0.001107	-0.046554	-0.485704	-1.733532	7.128299	0.000104	0.000066
	-0.000014	-0.000009	0.001046	-0.000211	0.000104	0.000005	0.000000
	0.000012	-0.000016	-0.000082	0.001946	0.000066	0.000000	0.000013

Table 2.27 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 6 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.117391	0.136923	1.094700	1.392565	-0.728766	0.015087	0.025222
Estimation Bias	0.017391	0.016923	0.844700	0.992565	-0.478766	0.000087	0.000222
Relative Bias (%)	17.391061	14.102760	337.880001	248.141164	191.506467	0.578843	0.889462
Mean Square Error	0.017537	0.015264	1.364450	1.821613	1.051643	0.000001	0.000004
Estimated Variance	0.017234	0.008417	-0.020849	0.005971	0.003093	-3.0e-06	-0.000001
	0.008417	0.014978	-0.007274	-0.015161	-0.004803	-4.0e-06	-0.000007
	-0.020849	-0.007274	0.650932	-0.074235	-0.052363	8.0e-05	-0.000017
	0.005971	-0.015161	-0.074235	0.836429	-0.167118	-1.9e-05	0.000259
	0.003093	-0.004803	-0.052363	-0.167118	0.822426	2.7e-05	-0.000069
	-0.000003	-0.000004	0.000080	-0.000019	0.000027	1.0e-06	0.000000
	-0.000001	-0.000007	-0.000017	0.000259	-0.000069	0.0e+00	0.000004

Table 2.28 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 10 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.107905	0.128697	0.739068	0.975283	-0.520053	0.015077	0.025153
Estimation Bias	0.007905	0.008697	0.489068	0.575283	-0.270053	0.000077	0.000153
Relative Bias (%)	7.905481	7.247586	195.627019	143.820738	108.021100	0.510277	0.612846
Mean Square Error	0.005027	0.006014	0.483225	0.652264	0.399502	0.000001	0.000002
Estimated Variance	0.004964	0.002694	-0.008104	0.002232	0.000865	-1e-06	3.0e-06
	0.002694	0.005938	-0.003291	-0.007385	-0.002469	-1e-06	2.0e-06
	-0.008104	-0.003291	0.244038	-0.016784	-0.016390	4e-05	0.0e+00
	0.002232	-0.007385	-0.016784	0.321314	-0.061896	-1e-06	9.9e-05
	0.000865	-0.002469	-0.016390	-0.061896	0.326574	6e-06	-4.5e-05
	-0.000001	-0.000001	0.000040	-0.000001	0.000006	1e-06	0.0e+00
	0.000003	0.000002	0.000000	0.000099	-0.000045	0e+00	2.0e-06

Table 2.29 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 20 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.103635	0.123359	0.480057	0.673107	-0.368339	0.015056	0.025098
Estimation Bias	0.003635	0.003359	0.230057	0.273107	-0.118339	0.000056	0.000098
Relative Bias (%)	3.635311	2.799248	92.022698	68.276807	47.335782	0.375099	0.391860
Mean Square Error	0.001945	0.001729	0.125470	0.171127	0.119301	0.000000	0.000001
Estimated Variance	0.001932	0.001101	-0.003401	0.000777	0.000052	0e+00	0.0e+00
	0.001101	0.001718	-0.002034	-0.002987	-0.000858	0e+00	0.0e+00
	-0.003401	-0.002034	0.072544	-0.001960	-0.005944	1e-05	5.0e-06
	0.000777	-0.002987	-0.001960	0.096539	-0.023367	1e-06	2.5e-05
	0.000052	-0.000858	-0.005944	-0.023367	0.105297	1e-06	-8.0e-06
	0.000000	0.000000	0.000010	0.000001	0.000001	0e+00	0.0e+00
	0.000000	0.000000	0.000005	0.000025	-0.000008	0e+00	1.0e-06

Table 2.30 Estimation Results for Two Dimensional CIR Process Driven by Two Dimensional Independent Brownian Motion with 200 Year Weekly Sample, 5000 Simulations.

	α_1	α_2	κ_{11}	κ_{22}	κ_{12}	σ_{11}	σ_{22}
True Value	0.100	0.120	0.250	0.400	0.550	0.015	0.025
Estimated Value	0.100414	0.120308	0.269135	0.424444	-0.259099	0.015020	0.025047
Estimation Bias	0.000414	0.000308	0.019135	0.024444	-0.009099	0.000020	0.000047
Relative Bias (%)	0.413602	0.256440	7.654090	6.111035	3.639420	0.136326	0.186166
Mean Square Error	0.000118	0.000135	0.003294	0.004787	0.005469	0.000000	0.000000
Estimated Variance	0.000118	0.000073	-0.000302	0.000010	-0.000012	0e+00	0e+00
	0.000073	0.000135	-0.000183	-0.000288	-0.000018	0e+00	0e+00
	-0.000302	-0.000183	0.002927	0.000002	0.000019	1e-06	0e+00
	0.000010	-0.000288	0.000002	0.004190	-0.001575	0e+00	2e-06
	-0.000012	-0.000018	0.000019	-0.001575	0.005386	0e+00	-1e-06
	0.000000	0.000000	0.000001	0.000000	0.000000	0e+00	0e+00
	0.000000	0.000000	0.000000	0.000002	-0.000001	0e+00	0e+00

samples sizes, the estimation effect is not as good as the O-U version. But we do see the same bias reduction effect when sample size increases.

2.5 Technical Proofs

2.5.1 Estimation under General Diffusion Process

Proof of Theorem 2.2.1 Since from (i), we know both $\phi_T(\eta)$ and $\phi_{Tr}(\eta)$ are uniformly bounded, say by B, and there exist C such that

$$|\phi_T(\eta) - \phi_{Tr}(\eta)| \leq C||\eta - \eta_0||^r$$

for all $\eta \in S$ and $n \geq N_0$. Therefore,

$$\begin{aligned}
& |\int_S |\phi_T(\eta)|^k f_T(\eta)d\eta - \int_S |\phi_{Tr}(\eta)|^k f_T(\eta)d\eta| \\
= & |\int_S k|\phi_T^{*1}(\eta)|^{k-1}|\phi_T(\eta) - \phi_{Tr}(\eta)| f_T(\eta)d\eta| \\
\leq & kCB^{B-1}\int_S ||\eta - \eta_0||^r f_T(\eta)d\eta \\
\leq & kCB^{k-1}\int_\Omega ||\eta - \eta_0||^r f_T(\eta)d\eta \\
\leq & kCB^{k-1}[\int_\Omega ||\eta - \eta_0||^R f_T(\eta)d\eta]^{r/R} \\
= & O(n^{\gamma r}),
\end{aligned} \tag{5.8}$$

by using Jensen's inequality and noting condition (ii), where $\phi_T^{*1}(\eta) \in (\phi_T(\eta) \wedge \phi_{Tr}(\eta), \phi_T(\eta) \vee \phi_{Tr}(\eta))$.

Now, we explore the boundary properties of $|\phi_{T_n}(\eta)|^k f_T(\eta)$ when $\eta \in S^c$. Because from (i), we also know $|\phi_{Tr}(\eta)|^k/||\eta - \eta_0||^k$ is bounded if $\eta \notin S$, and the bound D is uniform for all $T \geq T_0$. Hence,

$$\int_{S^c} |\phi_{T_n}(\eta)|^k f_T(\eta)d\eta \leq D\int_{S^c} ||\eta - \eta_0||^R f(\eta)d\eta = O(n^{-\gamma r}). \tag{5.9}$$

At the same time, if we denote C as a lower bound of $||\eta-\eta_0||$ on S^c, then by Hölder inequality,

$$\begin{aligned}
& \int_{S^c} |\phi_T(\eta)|^k f_T(\eta) d\eta = \int_{S^c} |\phi_T(\eta)|^k f(\eta)^{\frac{K-k}{K}} d\eta \\
\leq\ & [\int_{S^c} |\phi_T(\eta)|^k f(\eta) d\eta]^{k/K} [\int_{S^c} f(\eta) d\eta]^{\frac{K-k}{K}} \\
\leq\ & O(n^{\frac{\lambda k}{K}}) O(n^{-\gamma R(K=k)/K}) C^{-R} \int_{S^c} ||\eta=\eta_0||^R f_T(\eta) d\eta \\
=\ & (O(n^{-\gamma R}))^{\frac{k}{K}} (O(n^{\gamma R}))^{\frac{K-k}{K}} \\
=\ & O(n^{\gamma R}). \qquad (5.10)
\end{aligned}$$

Therefore,

$$E(|\phi_T(\hat{\eta})|^k) - E(|\phi_{T_r}(\hat{\eta})|^k) = O(n^{\gamma r}),$$

by combining (1), (2) and (3). □

Proof of Theorem 2.2.3 Now, let's first do the Taylor expansion to

$$\frac{1}{n}\sum_{i=1}^{n}[U \cdot g(X_i, X_{i+1}; \theta)]^j$$

and prove that the expectation of the remainder terms is of order n^{-2}.

Because

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}[U \cdot g(X_i, X_{i+1}; \theta)]^j = \frac{1}{n}\sum_{i=1}^{n} f_i^j(\eta) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Bigg\{ f^j(X_i, X_{i+1}; \eta_0) + \sum_k \frac{\partial f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^k}(\eta^k-\eta_0^k) \\
&\qquad + \frac{1}{2}\sum_{k,l}\frac{\partial^2 f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^l \partial \eta^k}(\eta^k-\eta_0^k)(\eta^l-\eta_0^l) \\
&\qquad + \frac{1}{6}\sum_{k,l,m}\frac{\partial^3 f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^m \partial \eta^l \partial \eta^k}(\eta^k-\eta_0^k)(\eta^l-\eta_0^l)(\eta^m-\eta_0^m) + O_p(n^{-2})\Bigg\} \\
&= (B^j+\beta^j) + (B^{j,k}+\beta^{j,k})(\eta^k-\eta_0^k) + \frac{1}{2}(B^{j,kl}+\beta^{j,kl})(\eta^k-\eta_0^k)(\eta^l-\eta_0^l) \\
&\quad + \frac{1}{6}(B^{j,klm}+\beta^{j,klm})(\eta^k-\eta_0^k)(\eta^l-\eta_0^l)(\eta^m-\eta_0^m) + O_p(n^{-2}) \qquad (5.11)
\end{aligned}$$

We proved 2.2.1 so that we can declare that the smaller order terms in equation (5.11) are still smaller order terms after we take expectation and hence we can estimate the bias and variance of certain parameters by taking expectation and variance for those main terms.

Now, we want to verify that

$$\eta^j - \eta_0^j = O_p(-B^j) = O_p(n^{-1/2}).$$

Note that $\beta^{j,k} = \triangle_j(k) = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$, using the tensor method, based on the previous Taylor expansion, we have

$$\begin{aligned}
& \eta^j - \eta_0^j \\
= & -B^j - B^{j,k}\{-B^k - B^{k,l}(\eta^l - \eta_0^l) - \frac{1}{2}(B^{k,lm} + \beta^{k,lm})(\eta^l - \eta_0^l)(\eta^m - \eta_0^m) - \zeta^j\} \\
& -\frac{1}{2}(B^{j,kl} + \beta^{j,kl})\{-B^k - B^{k,m}(\eta^m - \eta_0^m - \frac{1}{2}(B^{k,mn} + \beta^{k,mn})(\eta^m - \eta_0^m)(\eta^n - \eta_0^n) - \zeta^k\} \\
& \{-B^l - B^{l,m}(\eta^m - \eta^m 0) - \frac{1}{2}(B^{l,mn} + \beta^{l,mn})(\eta^m - \eta_0^m)(\eta^n - \eta_0^n) + \zeta^l\} - \zeta^j \qquad (5.12)
\end{aligned}$$

$$\begin{aligned}
= & -B^j - B^{j,k}\{-B^k - B^{k,l}(\eta^l - \eta_0^l) - \frac{1}{2}(B^{k,lm} + \beta^{k,lm})(\eta^l - \eta_0^l)(\eta^m - \eta_0^m)\} \\
& -\frac{1}{2}(B^{j,kl} + \beta^{j,kl})\{-B^k - B^{k,m}(\eta^m - \eta_0^m) - \frac{1}{2}(B^{k,mn} + \beta^{k,mn})(\eta^m - \eta_0^m)(\eta^n - \eta_0^n)\} \\
& \{-B^l - B^{l,m}(\eta^m - \eta_0^m) - \frac{1}{2}(B^{l,mn} + \beta^{l,mn})(\eta^m - \eta_0^m)(\eta^n - \eta_0^n)\} + \text{remainder terms} \\
= & -B^j + B^{j,k}B^k - \frac{1}{2}\beta^{j,kl}B^k B^l + O_p(n^{-3/2}). \qquad (5.13)
\end{aligned}$$

The order holds because $B^j = O_p(n^{-1/2})$ (Here, we are not taking expectation yet, so the order is ensured by $B^j = O_p(n^{-1/2})$ rather than Theorem 1.), which is shown below. Note that X_i is stable implies that $b_i^{j;j_1,\ldots,j_k}$ is stable too.

Then, based on the central limit theory for $\alpha-$mixing processes, we know

$$\sigma^2_{j;j_1,\ldots,j_k} < \infty \text{ and } \lim_n P\Big(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n b_i^{j;j_1,\ldots,j_k} < z\Big) = \phi(z),$$

i.e. $\sqrt{n}B^{j;j_1,\ldots,j_k} \to N(0, \sigma^2_{j;j_1,\ldots,j_k})$. Now that we have established (5.13) and Theorem 1, we can take expectation on both side of (5.12) and find

$$E(\eta^j - \eta_0^j) \leq O(n^{-1/2}).$$

And hence, the main terms that we need to consider when deriving high order property of $(\eta - \eta_0)$ are only $-B^j + B^{j,k}B^k - \frac{1}{2}\beta^{j,kl}B^kB^l$. To be more specific, we only need to consider

$$E(B^{j,k}B^k - \frac{1}{2}\beta^{j,kl}B^kB^l) \text{ and } var(-B^j),$$

since $E(-B^j) = 0$.

Therefor, we found

$$\begin{aligned} var(\eta_j) &= var(-B^j) = E[(-\frac{1}{n}\sum_{i=1}^{n} f^j(X_i, X_i; \eta_0) + \varpi^j)^2] \\ &= \frac{1}{n}E[(-f^j(X_i, X_i; \eta_0) + \varpi^j)^2] = \frac{1}{n}\{E[(u_{jk}g^k(X_i, X_i; \eta_0))^2] + [E(u_{jk}g^k(X_i, X_i; \eta_0))]^2\}, \end{aligned}$$

and

$$\begin{aligned} & n \cdot bias(\eta^j) \\ = \; & nE[-B^j + B^{j,k}B^k] - \frac{n}{2}\beta^{j,kl}E[B^kB^l] \\ = \; & E[\frac{\partial f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^k} f^k(X_i, X_{i+1}; \eta_0)] + \frac{1}{n}\sum_{i=1}^{n-1}(n-i)E[\frac{\partial f^j(X_{i+1}, X_{i+2}; \eta_0)}{\partial \eta^k} f^k(X_1, X_2; \eta_0)] \\ & -\varpi^k E[\frac{\partial f^j(X_i, X_{i+1}; \eta_0)}{\partial \eta^k}] - \frac{1}{2}E[\frac{\partial^2 f^j(X_1, X_2; \eta_0)}{\partial \eta^k \partial \eta^l}]\{E[f^k(X_1, X_2; \eta_0) f^l(X_1, X_2; \eta_0)] - \varpi^k\varpi^l\} \\ = \; & E[u_{ji}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{ks}g^s(X_\alpha, X_{\alpha+1}; \theta_0)] \\ & -\frac{1}{2}E[u_{ji}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k \partial \theta^l}]\{E[u_{ks}g^s(X_\alpha, X_{\alpha+1}; \theta_0) \cdot u_{lm}g^m(X_\alpha, X_{\alpha+1}; \theta_0)] \\ & -E[u_{km}g^m(X_\alpha, X_{\alpha+1}; \theta_0)]E[u_{li}g^i(X_\alpha, X_{\alpha+1}; \theta_0)]\} \\ & +\sum_{\alpha=j}^{n-1}\frac{n-\alpha}{n}E[u_{ji}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{ks}g^s(X_1, X_2; \theta_0)] - E[u_{kl}g^l(X_\alpha, X_{\alpha+1}; \theta_0)]E[u_{jm}\frac{g^m(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k}]. \end{aligned}$$

Transforming back to $\theta = V\eta$, where $V := (v_{ij})_{p\times p}$, we arrived at our conclusion that

$$\begin{aligned}
bias(\hat{\theta}_m) &= \sum_{j=1}^{p} v_{mj} bias(\hat{\eta}^j) \\
&= \frac{1}{n}\sum_{j=1}^{p}\sum_{b=0}^{B}\frac{\triangle^b}{b!} v_{mj}(\gamma^{b(j,k;k)} - \frac{1}{2}\beta^{j,kl}\gamma^{b(k;l)}) + o(\triangle^b), \\
var(\hat{\theta}^m) &= \sum_{j=1}^{p}\sum_{k=1}^{p} v_{mj}v_{mk} cov(\hat{\eta}_j, \hat{\eta}_k) \\
&= \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{b=0}^{B}\frac{1}{n}\frac{\triangle^b}{b!} v_{mj}v_{mk}\gamma^{b(j;k)}.
\end{aligned}$$

We want to comment that as proved in Genon-Catalot, Jeantheau and Laredo (2000), many popular financial models like the Vasicek model and the CIR model all satisfy the strong mixing condition we assumed here.

2.5.2 Estimation under One Dimensional O-U Process

We pursue the final bias and variance form for parameters in one dimensional O-U processes in five steps. First, we specify the estimation equation. Second, to find out the transforming matrix U, we work out the first order derivatives of the estimating equations with respect to all parameters, take expectation, and get the inverse matrix. Third, we work out the second order derivatives of the estimating equations with respect to all parameters. Fourth, we calculate these needed moments through either infinitesimal generator method or transitional density, since it is known for the one dimensional O-U process. Finally, we plug all those specific values into the forms we derived for bias and variance for a general process and simplify the results.

Step 1, we propose an estimating equation. Here, we adopt the maximal likelihood equation as our estimating equation. The only purpose of choosing this rather than other fancy ones is to facilitate comparison of our result with other existing ones. However, unlike others, we are not going to solve this equation for a specific form of the estimator, which is possible for one dimensional O-U process but not for most other diffusion processes.

The likelihood function is

$$\begin{aligned} L &= \log\prod_{i=1}^{n}\{2\pi\cdot\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\triangle})\}^{-\frac{1}{2}}\exp\{-\frac{[X_i-X_{i-1}e^{-\kappa\triangle}-\alpha(1-e^{-\kappa\triangle})]^2}{2\cdot\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\triangle})}\} \\ &= \sum_{i=1}^{n}[\frac{1}{2}\log\kappa-\frac{1}{2}\log\pi-\log(\sigma^2)-\frac{1}{2}\log(1-e^{-2\kappa\triangle})-\frac{(X_i-\alpha-e^{-\kappa\triangle(X_{i-1}-\alpha)})^2}{\sigma^2\kappa^{-1}(1-e^{-2\kappa\triangle})}]. \end{aligned}$$

Denote

$$\begin{aligned} \chi_i &= X_i-X_{i-1}e^{-\kappa\triangle}-\alpha(1-e^{-\kappa\triangle}) = X_i-\alpha-e^{-\kappa\triangle}(X_{i-1}-\alpha), \\ \beta_{i-1} &= X_{i-1}-\alpha, \end{aligned}$$

and observe that

$$\begin{aligned} \frac{\partial\chi_i}{\partial\kappa} &= \triangle e^{-\kappa\triangle}(X_{i-1}-\alpha) = \triangle e^{-\kappa}\beta_{i-1}, \\ \frac{\partial\chi_i}{\partial\alpha} &= -(1-e^{\kappa\triangle}), \\ \frac{\partial^2\chi_i}{\partial\kappa^2} &= -\triangle 2e^{-\kappa\triangle}(X_{i-1}-\alpha) = -\triangle^2 e^{-\kappa\triangle}\beta_{i-1}, \\ \frac{\partial^2\chi_i}{\partial\alpha^2} &= 0, \\ \frac{\partial^2\chi_i}{\partial\kappa\partial\alpha} &= -\triangle e^{-\kappa\triangle}, \end{aligned}$$

$$\begin{aligned} \frac{\partial\beta_{i-1}}{\partial\kappa} &= 0, \\ \frac{\partial\beta_{i-1}}{\partial\alpha} &= -1, \\ \frac{\partial\frac{1}{\sigma^2}}{\partial\sigma^2} &= -\frac{1}{\sigma^4}, \\ \frac{\partial^2\frac{1}{\sigma^2}}{\partial(\sigma^2)^2} &= \frac{2}{\sigma^6}. \end{aligned}$$

We will use the following three equations as our estimating equations:

$$\begin{aligned} g^1 &= \frac{\partial L}{\partial\kappa} = \frac{1}{2\kappa}-\frac{\triangle e^{-2\kappa\triangle}}{1-e^{-2\kappa\triangle}}-\frac{2\chi_i\triangle e^{-\kappa\triangle}\beta_{i-1}}{\kappa^{-1}\sigma^2(1-e^{-2\kappa\triangle})}+\frac{\chi_i^2[-(1-e^{-2\kappa\triangle})+2\kappa\triangle e^{-2\kappa\triangle}]}{\sigma^2(1-e^{-2\kappa\triangle})^2} = 0, \\ g^2 &= \frac{\partial L}{\partial\alpha} = \frac{2\chi_i}{\kappa^{-1}\sigma^2(1+e^{-\kappa\triangle})} = 0, \\ g^3 &= \frac{\partial L}{\partial(\sigma^2)} = -\frac{1}{2\sigma^2}+\frac{\chi_i^2}{\sigma^4\kappa^{-1}(1-e^{-2\kappa\triangle})} = 0. \end{aligned}$$

In order to work out the value of $bias(\kappa)$, note

$$
\begin{aligned}
& n \cdot bias(\kappa) \\
= \;& n \cdot bias(\eta^1) \\
= \;& nE[B^{1,k}B^k] - \frac{n}{2}\beta^{1,kl}E[B^kB^l] \\
= \;& E\Big[\frac{\partial f^1(X_i, X_{i+1}; \eta_0)}{\partial \eta^k} f^k(X_i, X_{i+1}; \eta_0)\Big] + \frac{1}{n}\sum_{i=1}^{n-1}(n-i)E\Big[\frac{\partial f^1(X_{i+1}, X_{i+2}; \eta_0)}{\partial \eta^k} f^k(X_1, X_2; \eta_0)\Big] \\
& -\frac{1}{2}E\Big[\frac{\partial^2 f^1(X_1, X_2; \eta_0)}{\partial \eta^k \partial \eta^l}\Big]E[f^k(X_1, X_2; \eta_0) f^l(X_1, X_2; \eta_0)] \\
= \;& E\Big[u_{1i}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{kj} g^j(X_\alpha, X_{\alpha+1}; \theta_0)\Big] \\
& -\frac{1}{2}E\Big[u_{1i}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k \partial \theta^l}\Big]E[u_{ki} g^i(X_\alpha, X_{\alpha+1}; \theta_0) \cdot u_{lj} g^j(X_\alpha, X_{\alpha+1}; \theta_0)] \\
& + \sum_{\alpha=1}^{n-1}\frac{n-\alpha}{n}E\Big[u_{1i}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{kj} g^j(X_1, X_2; \theta_0)\Big].
\end{aligned}
$$

Here, $\theta = (\kappa, \alpha, \sigma^2)$.

First, we can work out that

$$
\begin{aligned}
\frac{\partial g^1}{\partial \kappa} &= \frac{\partial^2 L}{\partial \kappa^2} = -\frac{1}{2\kappa^2} + \frac{2\triangle^2 e^{-2\kappa\triangle}}{(1-e^{-2\kappa\triangle})^2} - \frac{2\triangle^2 e^{-2\kappa\triangle\beta_{i-1}^2}}{\kappa^{-1}\sigma^2(1-e^{-2\kappa\triangle})} \\
& \quad + \frac{-4\triangle e^{-\kappa\triangle}(1-e^{-2\kappa\triangle}) + 2\kappa\triangle^2 e^{-\kappa\triangle(1+3e^{-2\kappa\triangle})}}{\sigma^2(1-e^{-2\kappa\triangle})^2}\chi_i\beta_{i-1} \\
& \quad + \frac{-4\kappa\triangle^2 e^{-2\kappa\triangle}(1+e^{-2\kappa\triangle}) + 4\triangle e^{-2\kappa\triangle}(1-e^{-2\kappa\triangle})}{\sigma^2(1-e^{-2\kappa\triangle})^3}\chi_i^2, \\
\frac{\partial g^2}{\partial \kappa} &= \frac{\partial^2 L}{\partial \kappa \partial \alpha} = \frac{2\kappa\triangle e^{-\kappa\triangle}}{\sigma^2(1+e^{-\kappa\triangle})}\beta_{i-1} + \frac{2(1+e^{-\kappa\triangle+\kappa\triangle^{-\kappa\triangle}})}{\sigma^2(1+e^{-\kappa\triangle})^2}\chi_i, \\
\frac{\partial g^3}{\partial \kappa} &= \frac{\partial^2 L}{\partial \kappa \partial(\sigma^2)} = \frac{2\kappa\triangle e^{-\kappa\triangle}}{\sigma^4(1-e^{-2\kappa\triangle})}\chi_i\beta_{i-1} + \frac{(1-e^{-2\kappa\triangle}) - 2\kappa\triangle e^{-2\kappa\triangle}}{\sigma^4(1-e^{-2\kappa\triangle})^2}\chi_i^2,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial g^1}{\partial \alpha} &= \frac{\partial^2 L}{\partial \kappa \partial \alpha} = \frac{\partial g^2}{\partial \kappa}, \\
\frac{\partial g^2}{\partial \alpha} &= \frac{\partial^2 L}{\partial \alpha^2} = -\frac{2\kappa(1-e^{-\kappa\triangle})}{\sigma^2(1+e^{-\kappa\triangle})}, \\
\frac{\partial g^3}{\partial \alpha} &= \frac{\partial^2 L}{\partial \alpha \partial(\sigma^2)} = -\frac{2\kappa\chi_i}{\sigma^4(1+e^{-\kappa\triangle})},
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial g^1}{\partial(\sigma^2)} &= \frac{\partial^2 L}{\partial\kappa\partial(\sigma^2)} = \frac{\partial g^3}{\partial\kappa},\\
\frac{\partial g^2}{\partial(\sigma^2)} &= \frac{\partial^2 L}{\partial\alpha\partial(\sigma^2)} = -\frac{2\kappa\chi_i}{\sigma^4(1+e^{-\kappa\triangle})},\\
\frac{\partial g^3}{\partial(\sigma^2)} &= \frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{2\kappa\chi_i^2}{\sigma^6(1-e^{-2\kappa\triangle})}.
\end{aligned}$$

So we have

$$E(\frac{\partial g}{\partial\theta}) = \begin{pmatrix} -\frac{1}{2\kappa^2} + \frac{2\kappa^{-1}\triangle e^{-2\kappa\triangle}}{1-e^{-2\kappa\triangle}} - \frac{\triangle^2 e^{-2\kappa\triangle}(1+e^{-2\kappa\triangle})}{(1-e^{-2\kappa\triangle})^2} & 0 & \frac{1}{2\kappa\sigma^2} - \frac{\triangle e^{-2\kappa\triangle}}{\sigma^2(1-e^{-2\kappa\triangle})} \\ 0 & -\frac{2\kappa(1-e^{-\kappa\triangle})}{\sigma^2(1+e^{-\kappa\triangle})} & 0 \\ \frac{1}{2\kappa\sigma^2} - \frac{\triangle e^{-2\kappa\triangle}}{\sigma^2(1-e^{-2\kappa\triangle})} & 0 & -\frac{1}{2\sigma^4} \end{pmatrix}.$$

Therefore,

$$det\Big[E(\frac{\partial g}{\partial\theta})\Big] = -\frac{\kappa\triangle^2 e^{-2\kappa\triangle}}{\sigma^6(1+e^{-\kappa\triangle})^2},$$

and

$$U = \Big[E(\frac{\partial g}{\partial\theta})\Big]^{-1} = \begin{pmatrix} -\frac{1-e^{-2\kappa\triangle}}{\triangle^2 e^{-2\kappa\triangle}} & 0 & -\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2\kappa\triangle}} + \frac{2\sigma^2}{\triangle} \\ 0 & -\frac{\sigma^2(1+e^{-\kappa\triangle})}{2\kappa(1-e^{-\kappa\triangle})} & 0 \\ -\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2\kappa\triangle}} + \frac{2\sigma^2}{\triangle} & 0 & -\frac{\sigma^4(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} + \frac{4\sigma^4}{\kappa\triangle} - \frac{2\sigma^4(1+e^{-2\kappa\triangle})}{1-e^{-2\kappa\triangle}} \end{pmatrix}.$$

Next, we can derive that

$$
\begin{aligned}
\frac{\partial^2 g^1}{\partial \kappa^2} &= \frac{\partial^3 L}{\partial \kappa^3} = \frac{1}{\kappa^3} - \frac{4\delta^3 e^{-2\kappa\triangle}(1+e^{-2\kappa\triangle})}{(1-e^{-2\kappa\triangle})^3} \\
&+ [\frac{4\kappa\triangle^3 e^{-2\kappa\triangle} - 6\triangle^2 e^{-2\kappa\triangle}}{\sigma^2(1-e^{-2\kappa\triangle})} + \frac{2\kappa\triangle^3 e^{-2\kappa\triangle}(1+5e^{-2\kappa\triangle})}{\sigma^2(1-e^{-2\kappa\triangle})^2}]\beta_{i-1}^2 \\
&+ \frac{6\triangle^2 e^{-\kappa\triangle}(1-e^{-2\kappa\triangle})(1+3e^{-2\kappa\triangle}) - \kappa\triangle^3 e^{-\kappa\triangle(2+32e^{-2\kappa\triangle}+14e^{-4\kappa\triangle})}}{\sigma^2(1-e^{-2\kappa\triangle})^3}\chi_i\beta_{i-1} \\
&+ \frac{-12\triangle^2 e^{-2\kappa\triangle}(1-e^{-4\kappa\triangle}) + 8\kappa\triangle^3 e^{-2\kappa\triangle}(1+4e^{-2\kappa\triangle}+e^{-4\kappa\triangle})}{\sigma^2(1-e^{-2\kappa\triangle})^4}\chi_i^2, \\
\frac{\partial^2 g^1}{\partial \alpha^2} &= \frac{\partial^3 L}{\partial\kappa\partial\alpha^2}, = \frac{-4\kappa\triangle e^{-\kappa\triangle} - 2(1-e^{-2}\kappa\triangle)}{\sigma^2(1+e^{-\kappa\triangle})^2} \\
\frac{\partial^2 g^1}{\partial(\sigma^2)^2} &= \frac{\partial^3 L}{\partial\kappa\partial(\sigma^2)^2} = \frac{-4\kappa\triangle e^{-\kappa\triangle}}{\sigma^6(1-e^{-2\kappa\triangle})}\chi_i\beta_{i-1} - \frac{2(1-e^{-2\kappa\triangle}) - 4\kappa\triangle e^{-2\kappa\triangle}}{\sigma^6(1-e^{-2\kappa\triangle})^2}\chi_i^2, \\
\frac{\partial^2 g^1}{\partial\kappa\partial\alpha} &= \frac{\partial^3 L}{\partial\kappa^2\partial\alpha} = \frac{4\triangle e^{-\kappa\triangle}(1+e^{-\kappa\triangle}) - 2\kappa\triangle^2 e^{-\kappa\triangle}(1-e^{-\kappa\triangle})}{\sigma^2(1+e^{-\kappa\triangle})^2}\beta_{i-1} \\
&+ \frac{-2\kappa\triangle^2 e^{-\kappa\triangle}(1-e^{-\kappa\triangle}) + 4\triangle e^{-\kappa\triangle(1+e^{-\kappa\triangle})}}{\sigma^2(1+e^{-\kappa\triangle})^3}\chi_i, \\
\frac{\partial^2 g^1}{\partial\kappa\partial(\sigma^2)} &= \frac{\partial^3 L}{\partial\kappa^2\partial(\sigma^2)} = \frac{4\triangle e^{-\kappa\triangle}(1-e^{-2\kappa\triangle}) - 2\kappa\triangle^2 e^{-\kappa\triangle}(1+3e^{-2\kappa\triangle})}{\sigma^4(1-e^{-2}\kappa\triangle)^2}\chi_i\beta_{i-1} \\
&+ \frac{2\kappa\triangle^2 e^{-2\kappa\triangle}}{\sigma^4(1-e^{-2\kappa\triangle})}\beta_{i-1}^2 + \frac{4\kappa\triangle^2 e^{-2\kappa\triangle}(1+e^{-2\kappa\triangle}) - 4\triangle e^{-2\kappa\triangle}(1-e^{-2\kappa\triangle})}{\sigma^4(1-e^{-2\kappa\triangle})^3}\chi_i^2, \\
\frac{\partial^2 g^1}{\partial\alpha\partial(\sigma^2)} &= \frac{\partial^3 L}{\partial\kappa\partial\alpha\partial(\sigma^2)} = -\frac{2\kappa\triangle e^{-\kappa\triangle}}{\sigma^4(1+e^{-\kappa\triangle})}\beta_{i-1} - \frac{2[1+e^{-\kappa\triangle}+\kappa\triangle e^{-\kappa\triangle}]}{\sigma^4(1+e^{-\kappa\triangle})^2}\chi_i,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 g^2}{\partial\kappa^2} &= \frac{\partial^3 L}{\partial\kappa^2\partial\alpha} = \frac{\partial^2 g^1}{\partial\kappa\partial\alpha}, \\
\frac{\partial^2 g^2}{\partial\alpha^2} &= \frac{\partial^3 L}{\partial\alpha^3} = 0, \\
\frac{\partial^2 g^2}{\partial(\sigma^2)^2} &= \frac{\partial^3 L}{\partial\alpha\partial(\sigma^2)^2} = \frac{4\kappa}{\sigma^6(1+e^{-\kappa\triangle})}\chi_i, \\
\frac{\partial^2 g^2}{\partial\kappa\partial\alpha} &= \frac{\partial^3 L}{\partial\kappa\partial\alpha^2} = \frac{\partial^2 g^1}{\partial\alpha^2}, \\
\frac{\partial^2 g^2}{\partial\kappa\partial(\sigma^2)} &= \frac{\partial^3 L}{\partial\kappa\partial\alpha\partial(\sigma^2)} = -\frac{2\kappa\triangle e^{-\kappa\triangle}}{\sigma^4(1+e^{-\kappa\triangle})}\beta_{i-1} - \frac{2[1+e^{-\kappa\triangle}+\kappa\triangle e^{-\kappa\triangle}]}{\sigma^4(1+e^{-\kappa\triangle})^2}\chi_i, \\
\frac{\partial^2 g^2}{\partial\alpha\partial(\sigma^2)} &= \frac{\partial^3 L}{\partial\alpha^2\partial(\sigma^2)} = \frac{2\kappa(1-e^{-\kappa\triangle})}{\sigma^4(1+e^{-\kappa\triangle})},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 g^3}{\partial \kappa^2} &= \frac{\partial^3 L}{\partial \kappa^2 \partial(\sigma^2)} = \frac{\partial^2 g^1}{\partial \kappa \partial(\sigma^2)}, \\
\frac{\partial^2 g^3}{\partial \alpha^2} &= \frac{\partial^3 L}{\partial \alpha^2 \partial(\sigma^2)} = \frac{\partial^2 g^2}{\partial \alpha \partial(\sigma^2)}, \\
\frac{\partial^2 g^3}{\partial(\sigma^2)^2} &= \frac{\partial^3 L}{\partial(\sigma^2)^3} = -\frac{1}{\sigma^6} + \frac{6\kappa}{\sigma^8(1-e^{-2\kappa\triangle})}\chi_i^2, \\
\frac{\partial^2 g^3}{\partial \kappa \partial \alpha} &= \frac{\partial^3 L}{\partial \kappa \partial \alpha \partial(\sigma^2)} = \frac{\partial^2 g^1}{\partial \alpha \partial(\sigma^2)}, \\
\frac{\partial^2 g^3}{\partial \kappa \partial(\sigma^2)} &= \frac{\partial^3 L}{\partial \kappa \partial(\sigma^2)^2} = \frac{\partial^2 g^1}{\partial(\sigma^2)^2}, \\
\frac{\partial^2 g^3}{\partial \alpha \partial(\sigma^2)} &= \frac{\partial^3 L}{\partial \alpha \partial(\sigma^2)^2} = \frac{\partial^2 g^2}{\partial(\sigma^2)^2}.
\end{aligned}
$$

Further, we find out

$$
\begin{aligned}
E[\beta_{\zeta+1}^2\beta_2\beta_1] &= \frac{\sigma^4 e^{-\kappa\triangle}}{4\kappa^2}(1-e^{-2(\zeta-1)\kappa\triangle}) + \frac{3\sigma^4}{4\kappa 2}e^{-(2\zeta-1)\kappa\triangle}, \\
E[\beta_{\zeta+1}^2\beta_1] &= \frac{\sigma^4}{4\kappa^2}(1+2e^{-2\zeta\kappa\triangle}), \\
E[\beta_{\zeta+1}^2\chi_2\beta_1] &= \frac{\sigma^4}{2\kappa^2}\{e^{-(2\zeta-1)\kappa\triangle} - e^{-(2\zeta+1)\kappa\triangle}\}, \\
E[\beta_{\zeta+1}\alpha_2] &= \frac{1}{2}\kappa^{-1}\sigma^2(e^{-(\zeta-1)\kappa\triangle} - e^{-(\zeta+1)\kappa\triangle}), \\
E[\chi_{\zeta+2}^2\chi_2^2] &= \frac{\sigma^4}{4\kappa}(1-e^{-2\kappa\triangle})^2, \\
E[\chi_{\zeta+1}\beta_{\zeta+1}\chi_2\beta_1] &= E[\chi_{\zeta+2}^2\chi_2\beta_1] = E[\chi_{\zeta+2}\chi_2] = E[\chi_{\zeta+2}\beta_{\zeta+1}\chi_2^2] = 0.
\end{aligned}
$$

Therefore, after some more tedious calculation, we have

$$
\begin{aligned}
& n \cdot bias(\kappa) \\
= \; & (\frac{1-e^{-2\kappa\triangle}}{\kappa\triangle^2 e^{-2\kappa\triangle}} - \frac{e^{\kappa\triangle}+e^{-\kappa\triangle}}{\triangle e^{-\kappa\triangle}}) + \frac{1+e^{\kappa\triangle}}{\triangle} + (\frac{-2(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2\kappa\triangle}} + \frac{4}{\triangle}) \\
& + [\frac{4}{\triangle(1-e^{-2\kappa\triangle})} - \frac{2}{\kappa\triangle e^{-2\kappa\triangle}} + \frac{3+e^{-2\kappa\triangle}}{2\triangle e^{-2\kappa\triangle}} + \frac{(1-e^{-2\kappa\triangle})^2}{2\kappa^3\triangle^4 e{-4\kappa\triangle}} - \frac{1-e^{-2\kappa\triangle}}{\kappa 2\triangle^3 e^{-2\kappa\triangle}} - \frac{1-e^{-2\kappa\triangle}}{\kappa\triangle^2 e^{-2\kappa\triangle}}] \\
& + (-\frac{1+e^{\kappa\triangle}}{\triangle}) + [-\frac{2}{\triangle} + \frac{6}{\kappa\triangle^2} - \frac{3(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^3 e^{-2\kappa\triangle}} - \frac{4e^{-2\kappa\triangle}}{\triangle(1-e^{-2\kappa\triangle})} + \frac{1-e^{-2\kappa\triangle}}{\kappa\triangle^2 e^{-2\kappa\triangle}} + \frac{(1-e^{-2\kappa\triangle})^2}{2\kappa^3\triangle^4}e^{-4\kappa\triangle}] \\
& + 0 + [\frac{3(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2\kappa\triangle}} - \frac{6}{\triangle} - \frac{(1-e^{-2\kappa\triangle})^2}{\kappa^3\triangle^4 e^{-4\kappa\triangle}} + \frac{4(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^3 e^{-2\kappa\triangle}} - \frac{4}{\kappa\triangle^2}] + 0 \\
& + 0 + \frac{2}{\triangle} + \frac{1+e^{-\kappa\triangle}}{\triangle e^{-\kappa\triangle}} + 0 + o(1) = \frac{4}{\triangle} + 2\kappa + o(1),
\end{aligned}
$$

i.e.,

$$bias(\hat{\kappa}) = \frac{4}{T} + \frac{2\kappa}{n} + o(\frac{1}{n}).$$

To appreciate how those expectations above are gained, we illustrate below that there are two ways for doing so.

I. Since the transition probability for the O-U process is known, we can use the conditional expectation to work it out directly.

$$\begin{aligned}
E[\beta_{\zeta+1}^2\beta_1^2] &= EE[[\beta_{\zeta+1}^2\beta_1^2|\mathcal{F}_\zeta]] = E[\beta_1^2 E[\beta_{\zeta+1}^2|\mathcal{F}_\zeta]] \\
&= E[\beta_1^2 E[\beta_{\zeta+1}^2|X_\zeta]] = E[\beta_1^2 E[(\chi_{\zeta+1} + e^{-\kappa\triangle}\beta_\zeta)^2|X_\zeta]] \\
&= E[\beta_1^2 E[\chi_{\zeta+1}^2 + 2e^{-\kappa\triangle}\chi_{\zeta+1}\beta_\zeta + e^{-\kappa\triangle}\beta_\zeta^2|X_\zeta]] \\
&= E[\beta_1^2 E[\frac{1}{2}\kappa^{-1}\sigma^2(1-e^{-2\kappa\triangle}) + e^{-2\kappa\triangle}\beta_\zeta^2]] \\
&= E[\beta_1^2 E[\frac{1}{2}\kappa^{-1}\sigma^2(1-e^{-2\kappa\triangle})(1+\ldots+e^{-2(\zeta-1)\kappa\triangle}) + e^{-2(\zeta-1)\kappa\triangle}\beta_1^4]] \\
&= \frac{1}{2}\kappa^{-1}\sigma^2(1-e^{-2\kappa\triangle})\frac{1-e^{-2\zeta\kappa\triangle}}{1-e^{-2\kappa\triangle}}(\frac{1}{2}\kappa^{-1}\sigma^2) + e^{-2\zeta\kappa\triangle}3(\frac{1}{2}\kappa^{-1}\sigma^2)^2 \\
&= \frac{\sigma^4}{4\kappa^2}(1+2e^{-2\zeta\kappa\triangle}).
\end{aligned}$$

II. Whenever we have the diffusion processes, we have the corresponding generators. We'll derive the value of the conditional expectations via the generator using the formula (1.5).

Because

$$E[\beta_{\zeta+1}^2\beta_1^2] = E[(X_{\zeta+1}-\alpha)^2(X_1-\alpha)^2] = E[(X_1-\alpha)^2 E[(X_{\zeta+1}-\alpha)^2|X_\zeta]],$$

we need to find the value of $E[(X_{\zeta+1}-\alpha)^2|X_\zeta]$. Let

$$f(x) = (x-\alpha)^2.$$

Since the generator is

$$Af(x) = \kappa(\alpha - x)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2},$$

we find that

$$\begin{aligned}
A_0 f(X_\zeta) &= (X_\zeta - \alpha)^2, \\
A_1 f(X_\zeta) &= -2\kappa(X_\zeta - \alpha)^2 + \sigma^2, \\
A_2 f(X_\zeta) &= (-2\kappa)^2(X_\zeta - \alpha)^2 + (-2\kappa)\sigma^2, \\
&\vdots \\
A_n f(X_\zeta) &= (-2\kappa)^n(X_\zeta - \alpha)^2 + (-2\kappa)^{n-1}\sigma^2,
\end{aligned}$$

Here, the subscript $i \in \{0,1,2,\ldots\}$ in A_i means to apply the infinitesimal generator A to $f(X_\zeta)$ i times. Hence,

$$\begin{aligned}
E[(X_{\zeta+1} - \alpha)^2 | X_\zeta] &= \sum_{n=0}^{\infty}(X_\zeta - \alpha)^2(-2\kappa)^n \triangle^n/n! + \sum_{n=1}^{\infty}\sigma^2(-2\kappa)^{n-1}\triangle^n/n! \quad (5.14) \\
&= (X_\zeta - \alpha)^2 e^{-2\kappa\triangle} - \frac{\sigma^2}{2\kappa}(e^{-2\kappa\triangle} - 1) \\
&= e^{-2\kappa\triangle}\beta_\zeta^2 + \frac{1}{2}\kappa^{-1}\sigma^2(1 - e^{-2\kappa\triangle}).
\end{aligned}$$

Finally, we have

$$E[\beta_{\zeta+1}^2\beta_1^2] = E[e^{-2\kappa\triangle}\beta_\zeta^2\beta_1^2 + \frac{1}{2}\kappa^{-1}\sigma^2(1 - e^{-2\kappa\triangle})\beta_1^2] = \frac{\sigma^4}{4\kappa^2}(1 + 2e^{-2\zeta\kappa\triangle}).$$

Similar to the work for $bias(\hat{\kappa})$, we derive

$$\begin{aligned}
& n \cdot bias(\sigma^2) \\
= \; & E\left[u_{3i}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{kj} g^j(X_\alpha, X_{\alpha+1}; \theta_0)\right] \\
& -\frac{1}{2}E\left[u_{3i}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k \partial \theta^l}\right] E[u_{ki} g^i(X_\alpha, X_{\alpha+1}; \theta_0) \cdot u_{lj} g^j(X_\alpha, X_{\alpha+1}; \theta_0)] \\
& + \sum_{\alpha=1}^{n-1}\frac{n-\alpha}{n}E\left[u_{3i}\frac{\partial g^i(X_\alpha, X_{\alpha+1}; \theta_0)}{\partial \theta^k} \cdot u_{kj} g^j(X_1, X_2; \theta_0)\right]
\end{aligned}$$

$$
\begin{aligned}
= \quad & [\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} - \frac{\sigma^2(1+3e^{-2\kappa\triangle})}{\kappa\triangle e^{-2\kappa\triangle}} + \frac{4\sigma^2 e^{-2\kappa\triangle}}{1-e^{-2\kappa\triangle}}] \\
& +[\frac{\sigma^2(1-e^{-\kappa\triangle})}{\kappa\triangle e^{-\kappa\triangle}} + \frac{2\sigma^2}{\kappa\triangle} - \frac{2\sigma^2 e^{-\kappa\triangle}}{1+e^{-\kappa\triangle}} - \frac{2\sigma^2(1+e^{-2\kappa\triangle})}{1-e^{-2\kappa\triangle}}] \\
& +[-\frac{2\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} + \frac{8\sigma^2}{\kappa\triangle}] - \frac{4\sigma^2(1+e^{-2\kappa\triangle})}{1-e^{-2\kappa\triangle}} \\
& +[\frac{\sigma^2(1-e^{-2\kappa\triangle})^2}{2\kappa^4\triangle^4 e^{-4\kappa\triangle}} - \frac{3\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} + \frac{\sigma^2(3+5e^{-2\kappa\triangle})}{2\kappa\triangle e^{-2\kappa\triangle}} + \frac{2\sigma^2(3+e^{-2\kappa\triangle})}{\kappa\triangle(1-e^{-2\kappa\triangle})} \\
& -\frac{2\sigma^2(1+e^{-2\kappa\triangle})}{1-e^{-2\kappa\triangle}} - \frac{2\sigma^2(-1+4e^{-2\kappa\triangle}+e^{-4\kappa\triangle})}{(1-e^{-2\kappa\triangle})^2} - \frac{2\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^3\triangle^3 e^{-2\kappa\triangle}}] \\
& +[-\frac{\sigma^2}{\kappa\triangle e^{-\kappa\triangle}} + \frac{\sigma^2(1+e^{-\kappa\triangle})}{1-e^{-\kappa\triangle}} - \frac{\sigma^2}{\kappa\triangle}] \\
& +[\frac{\sigma^2(3-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} + \frac{\sigma^2(1-e^{-2\kappa\triangle})^2}{2\kappa^4\triangle^4 e^{-4\kappa\triangle}} - \frac{4\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^3\triangle^3 e^{-2\kappa\triangle}} \\
& +\frac{10\sigma^2}{\kappa^2\triangle^2} - \frac{4\sigma^2}{\kappa\triangle} - \frac{8\sigma^2(1+e^{-2\kappa\triangle})}{\kappa\triangle(1-e^{-2\kappa\triangle})} + \frac{4\sigma^2}{1-e^{-2\kappa\triangle}} + \frac{8\sigma^2 e^{-2\kappa\triangle}}{(1-e^{-2\kappa\triangle})^2}] \\
& +0+[\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} - \frac{4\sigma^2}{\kappa\triangle} + \frac{4\sigma^2 e^{-2\kappa\triangle}}{1-e^{-2}\kappa\triangle} - \frac{12\sigma^2}{\kappa^2\triangle^2} \\
& +\frac{8\sigma^2 e^{-2\kappa\triangle}}{\kappa\triangle(1-e^{-2\kappa\triangle})} + \frac{6\sigma^2(1-e^{-2\kappa\triangle})}{\kappa^3\triangle^3 e^{-2\kappa\triangle}} - \frac{\sigma^2(1-e^{-2\kappa\triangle})^2}{\kappa^4\triangle^4 e^{-4\kappa\triangle}}]+0+0 \\
& +[\frac{\sigma^2(1+e^{-\kappa\triangle})}{n\kappa\triangle e^{-2\kappa\triangle}} - \frac{2\sigma 2}{n(1-e^{-\kappa\triangle})}]\cdot[(n-1)e^{-\kappa\triangle} - \frac{e^{-2\kappa\triangle}(1-e^{-(n-2)\kappa\triangle})}{1-e^{\kappa\triangle}} - e^{-n\kappa\triangle}]+0 \\
= \quad & -2\sigma^2+o(1),
\end{aligned}
$$

$$
\begin{aligned}
& n\cdot bias(\alpha) \\
= \quad & E\Big[u_{2i}\frac{\partial g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial\theta^k}\cdot u_{kj}g^j(X_\alpha, X_{\alpha+1};\theta_0)\Big] \\
& -\frac{1}{2}E\Big[u_{2i}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial\theta^k\partial\theta^l}\Big]E[u_{ki}g^i(X_\alpha, X_{\alpha+1};\theta_0)\cdot u_{lj}g^j(X_\alpha, X_{\alpha+1};\theta_0)] \\
& +\sum_{\alpha=1}^{n-1}\frac{n-\alpha}{n}E\Big[u_{2i}\frac{\partial g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial\theta^k}\cdot u_{kj}g^j(X_1, X_2;\theta_0)\Big] \\
= \quad & o(1),
\end{aligned}
$$

$$
\begin{aligned}
var(\kappa) \quad = \quad & \frac{1}{n}E[(u_{1i}g^i(X_\alpha, X_{\alpha+1};\theta_0))^2] \\
= \quad & \frac{1}{n}\frac{1-e^{-2\kappa\triangle}}{\triangle^2 e^{-2\kappa\triangle}} = \frac{2\kappa}{T}+o(\frac{1}{T}),
\end{aligned}
$$

$$\begin{aligned}
var(\alpha) &= \frac{1}{n}E[(u_{2i}g^i(X_\alpha, X_{\alpha+1};\theta_0))^2] \\
&= \frac{1}{n}\frac{\sigma^2}{2\kappa}\frac{1+e^{-\kappa\triangle}}{1-e^{-\kappa\triangle}} = \frac{\sigma^2}{T\kappa^2} + o(\frac{1}{T}), \\
var(\sigma^2) &= \frac{1}{n}E[(u_{3i}g^i(X_\alpha, X_{\alpha+1};\theta_0))^2] \\
&= \frac{1}{n}[2\sigma^4 + \frac{\sigma^4(1-e^{-2\kappa\triangle})}{\kappa^2\triangle^2 e^{-2\kappa\triangle}} - \frac{4\sigma^4}{\kappa\triangle} + \frac{4\sigma^4 e^{-2\kappa\triangle}}{1-e^{-2\kappa\triangle}}] \\
&= \frac{2\sigma^4}{n} + o(\frac{1}{n}), \\
cov(\hat{\kappa}, \hat{\sigma}^2) &= o(\frac{1}{n}), \\
cov(\hat{\kappa}, \hat{\alpha}) &= \frac{1}{n}E[(u_{11}g^1 + u_{13}g^3)(u_{31} + u_{33}g^3)] \\
& \quad \frac{1}{n}(\frac{\sigma^2(1-e^{-2\kappa\triangle})}{\kappa\triangle^2 e^{-2}\kappa\triangle} - \frac{2\sigma^2}{\triangle}) = \frac{2\sigma^2\kappa}{n}, \\
cov(\hat{\sigma}^2, \hat{\alpha}) &= o(\frac{1}{n}).
\end{aligned}$$

2.5.3 Estimation under Two Dimensional O-U Process

Technical details for estimation under two dimensional O-U process driven by independent Brownian motions are provided in this subsection. We adopt the two dimensional Ornstein-Uhlenbeck diffusion process

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t,$$

or write in the form

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}\begin{pmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix},$$

where $\kappa_{21} \neq 0$ (or it is a degenerated case) and $\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix}_{t\geq 0}$ is a standard two dimensional Brownian motion. Because

$$e^{\kappa t}dX_t = e^{\kappa t}\kappa(\alpha - X_t)dt + e^{\kappa t}\sigma dW_t,$$

$$d(e^{\kappa t}X_t) = \kappa e^{\kappa t}X_t dt + e^{\kappa t}dX_t = e^{\kappa t}\kappa\alpha dt + e^{\kappa t}\sigma dW_t,$$

$$e^{\kappa t}X_t - X_0 = \int_0^t e^{\kappa s}\kappa\alpha ds + \int_0^t e^{\kappa s}\sigma dW_s,$$

we see,

$$X_t \quad = \quad e^{-\kappa\triangle}X_{t-1} + (I - e^{-\kappa\triangle})\alpha + e^{-\kappa t\triangle}\int_{(t-1)\triangle}^{t\triangle} e^{\kappa s}\sigma dW_s.$$

When $\kappa_{11} \neq \kappa_{22}$, let $b = \frac{\kappa_{21}}{\kappa_{22}-\kappa_{11}}$, we see[5]

$$\exp\begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} = \exp\left\{\begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix}\begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}\right\}$$

$$= \begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix}\exp\begin{pmatrix} \kappa_{11}t & 0 \\ 0 & \kappa_{22}t \end{pmatrix}\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix}\begin{pmatrix} e^{\kappa_{11}} & 0 \\ 0 & e^{\kappa_{22}} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} = \begin{pmatrix} e^{\kappa_{11}} & 0 \\ b(e^{\kappa_{22}} - e^{\kappa_{11}}) & e^{\kappa_{22}} \end{pmatrix},$$

while when $\kappa_{11} = \kappa_{22}$,

$$\exp\begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} = \exp\left\{\kappa_{11}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \kappa_{21}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right\}$$

$$= \exp\left\{\kappa_{11}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}\exp\left\{\kappa_{21}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right\}$$

$$= \begin{pmatrix} e^{\kappa_{11}} & 0 \\ 0 & e^{\kappa_{22}} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \kappa_{21} & 1 \end{pmatrix} = \begin{pmatrix} e^{\kappa_{11}} & 0 \\ \kappa_{21}e^{\kappa_{11}} & e^{\kappa_{11}} \end{pmatrix}.$$

Let

$$U_i^0 = e^{-\kappa\triangle}X_{i-1} + (I - e^{-\kappa\triangle})\alpha.$$

---

[5] So $\exp\left\{-\begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}\triangle\right\} = \begin{pmatrix} e^{-\kappa_{11}\triangle} & 0 \\ b(e^{-\kappa_{22}\triangle} - e^{-\kappa_{11}\triangle}) & e^{-\kappa_{22}\triangle} \end{pmatrix}$.

We have when $\kappa_{11} \neq \kappa_{22}$,

$$\begin{aligned}
U_i &= X_i - U_i^0 = \begin{pmatrix} X_{i,1} - U_{i1}^0 \\ X_{i,2} - U_{i2}^0 \end{pmatrix}, \\
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1), \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{22}\triangle + \frac{\kappa_{22}^2}{2}\triangle^2 - \frac{\kappa_{22}^3}{6}\triangle^3 + \frac{\kappa_{22}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2), \\
&\quad -\kappa_{21}\triangle\{1 - \frac{1}{2}\triangle(\kappa_{11} + \kappa_{22}) + \frac{1}{6}\triangle^2(\kappa_{11}^2 + \kappa_{11}\kappa_{22} + \kappa_{22}^2) \\
&\quad -\frac{1}{24}\triangle^3(\kappa_{11}^2 + \kappa_{22}^2)(\kappa_{11} + \kappa_{22})\}(X_{i-1,1} - \alpha_1).
\end{aligned}$$

When $\kappa_{11} = \kappa_{22}$,

$$\begin{aligned}
U_{i1}^0 &= \alpha_1 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,1} - \alpha_1) \\
U_{i2}^0 &= \alpha_2 + (1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3 + \frac{\kappa_{11}^4}{24}\triangle^4)(X_{i-1,2} - \alpha_2) \\
&\quad -\kappa_{21}\triangle(1 - \kappa_{11}\triangle + \frac{\kappa_{11}^2}{2}\triangle^2 - \frac{\kappa_{11}^3}{6}\triangle^3)(X_{i-1,1} - \alpha_1)
\end{aligned}$$

Denote

$$V = var\Big(e^{-\kappa t\triangle}\int_{(t-1)\triangle}^{t\triangle} e^{\kappa s}\sigma dW_s\Big) = \int_{(t-1)\triangle}^{t\triangle} e^{\kappa(-t\triangle+s)}\sigma\sigma^T(e^{\kappa(-t\triangle+s)})^T ds.$$

Then we find when $\kappa_{11} \neq \kappa_{22}$, $V = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}$,

$$\begin{aligned}
v_{11} &= \sigma_{11}^2\frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}} \\
&= \sigma_{11}^2\triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5), \\
v_{12} &= v_{21} = b\sigma_{11}^2(\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} - \frac{1 - e^{-2\kappa_{11}\triangle}}{2\kappa_{11}}) \\
&= \sigma_{11}^2\kappa_{21}\triangle^2\{-\frac{1}{2} + \triangle(\frac{1}{2}\kappa_{11} + \frac{1}{6}\kappa_{22}) - \triangle^2(\frac{7}{24}\kappa_{11}^2 + \frac{1}{6}\kappa_{11}\kappa_{22} + \frac{1}{24}\kappa_{22}^2)\} + O(\triangle^5), \\
v_{22} &= b^2\sigma_{11}^2\{\frac{1 - e^{-2\kappa_{11}}}{2\kappa_{11}} - 2\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\triangle}}{\kappa_{11} + \kappa_{22}} + \frac{1 - e^{-2\kappa_{11}}}{2\kappa_{22}}\} + \sigma_{22}^2\frac{1 - e^{-2\kappa_{22}}}{\kappa_{22}} \\
&= \sigma_{11}^2\kappa_{21}^2\triangle^3\{\frac{1}{3} - \triangle\frac{1}{4}(\kappa_{11} + \kappa_{22})\} + \sigma_{22}^2\triangle(1 - \kappa_{22}\triangle + \frac{2}{3}\kappa_{22}^2\triangle^2 - \frac{1}{3}\kappa_{22}^3\triangle^3) + O(\triangle^5).
\end{aligned}$$

When $\kappa_{11} = \kappa_{22}$,

$$
\begin{aligned}
v_{11} &= \sigma_{11}^2 \triangle (1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5),\\
v_{12} &= v_{21} = \sigma_{11}^2 \kappa_{21}\triangle^2(-\frac{1}{2} + \triangle\frac{2}{3}\kappa_{11} - \triangle^2\frac{1}{2}\kappa_{11}^2) + O(\triangle^5),\\
v_{22} &= \sigma_{11}^2\kappa_{21}^2\triangle^3(\frac{1}{3} - \frac{1}{2}\kappa_{11}\triangle) + \sigma_{22}^2\triangle(1 - \kappa_{11}\triangle + \frac{2}{3}\kappa_{11}^2\triangle^2 - \frac{1}{3}\kappa_{11}^3\triangle^3) + O(\triangle^5).
\end{aligned}
$$

We observe the main orders for the conditional expectation and the variance of X_i given X_{i-1} when $\kappa_{11} = \kappa_{22}$ are the same as letting $\kappa_{11} = \kappa_{22}$ in the $\kappa_{11} \neq \kappa_{22}$ case. Therefore, we only need to derive those results for $\kappa_{11} \neq \kappa_{22}$ and then let $\kappa_{11} = \kappa_{22}$ to get the result for $\kappa_{11} = \kappa_{22}$.

In summary, we know the transitional distribution

$$X_t | X_{t-1} \sim N_2(U_i^0, V),$$

and the stationary distribution

$$X_t \sim N_2(\alpha, V_0),$$

where $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, V_0 = \begin{pmatrix} \frac{\sigma_{11}^2}{2\kappa_{11}} & b\sigma_{11}^2(\frac{1}{\kappa_{11}+\kappa_{22}} - \frac{1}{2\kappa_{11}}) \\ b\sigma_{11}^2(\frac{1}{\kappa_{11}+\kappa_{22}} - \frac{1}{2\kappa_{11}}) & b^2\sigma_{11}^2(\frac{1}{2\kappa_{11}} - \frac{2}{\kappa_{11}+\kappa_{22}} + \frac{1}{2\kappa_{22}}) + \frac{\sigma_{22}^2}{2\kappa_{22}} \end{pmatrix}$. Hence, the the log likelihood function is

$$l(\theta) = \sum_{i=1}^{n} \{-\log(2\pi) - \frac{1}{2} l^*(\theta)\},$$

where $\theta = \{\alpha_1, \alpha_2, \kappa_{11}, \kappa_{22}, \kappa_{21}, \sigma_{11}^2, \sigma_{22}^2\}$,

$$l^*(\theta) = \log|V| + U_i^T V^{-1} U_i.$$

And we adopt the estimating equation to be

$$\frac{\partial l^*(\theta)}{\partial \theta}.$$

That is, we adopt the estimating equations to be $h_i = \frac{\partial l^*(\theta)}{\partial \theta_i}$, $i = 1, ..., 7$, which are calculated with the aid of Mathematica and not provided here. Write

$$\begin{aligned}
v_{\alpha 11} &= \sigma_{11}^2 \frac{1 - e^{-2\kappa_{11}\alpha\triangle}}{2\kappa_{11}}, \\
v_{\alpha 21} &= \frac{\kappa_{21}}{\kappa_{22} - \kappa_{11}} \sigma_{11}^2 \left(\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\alpha\triangle}}{\kappa_{11} + \kappa_{22}} - \frac{1 - e^{-2\kappa_{11}\alpha\triangle}}{2\kappa_{11}} \right), \\
v_{\alpha 22} &= \left(\frac{\kappa_{21}}{\kappa_{22} - \kappa_{11}}\right)^2 \sigma_{11}^2 \left\{ \frac{1 - e^{-2\kappa_{11}\alpha\triangle}}{2\kappa_{11}} - 2\frac{1 - e^{-(\kappa_{11}+\kappa_{22})\alpha\triangle}}{\kappa_{11} + \kappa_{22}} + \frac{1 - e^{-2\kappa_{22}\alpha\triangle}}{2\kappa_{22}} \right\} + \sigma_{22}^2 \frac{1 - e^{-2\kappa_{22}\alpha\triangle}}{\kappa_{22}}.
\end{aligned}$$

Note:

$$\begin{aligned}
E[U_{i1}|X_{i-\alpha}] &= E[U_{i1}^3|X_{i-\alpha}] = 0, \\
E[U_{i1}^2|X_{i-\alpha}] &= v_{\alpha 11}, \\
E[U_{i1}^4|X_{i-\alpha}] &= 3v_{\alpha 11}^2, \\
E[U_{i2}|X_{i-\alpha}] &= E[U_{i2}^3|X_{i-\alpha}] = 0, \\
E[U_{i2}^2|X_{i-\alpha}] &= v_{\alpha 22}, \\
E[U_{i2}^4|X_{i-\alpha}] &= 3v_{\alpha 22}^2, \\
E[U_{i1}U_{i2}|X_{i-\alpha}] &= v_{\alpha 12}, \\
E[U_{i1}U_{i2}^2|X_{i-\alpha}] &= E[U_{i1}^2U_{i2}|X_{i-\alpha}] = 0, \\
E[U_{i1}^2U_{i2}^2|X_{i-\alpha}] &= 2v_{\alpha 12}^2 + v_{\alpha 11}v_{\alpha 22}, \\
E[U_{i1}U_{i2}^3|X_{i-\alpha}] &= 3v_{\alpha 21}v_{\alpha 22}, \\
E[U_{i1}^3U_{i2}|X_{i-\alpha}] &= 3v_{\alpha 21}v_{\alpha 11}, \\
E[X_{i1} - \alpha] &= E[(X_{i1} - \alpha)^3] = 0, \\
E[(X_{i1} - \alpha)^2] &= v0_{11}. \\
E[(X_{i1} - \alpha)^4] &= 3v0_{11}^2, \\
E[X_{i2} - \alpha] &= E[(X_{i2} - \alpha)^3] = 0, \\
E[(X_{i2} - \alpha)^2] &= v0_{22}, \\
E[(X_{i2} - \alpha)^4] &= 3v0_{22}^2,
\end{aligned}$$

$$
\begin{aligned}
E[(X_{i1}-\alpha)(X_{i2}-\alpha)] &= v0_{21}, \\
E[(X_{i1}-\alpha)^2(X_{i2}-\alpha)] &= E[(X_{i1}-\alpha)(X_{i2}-\alpha)^2] = 0, \\
E[(X_{i1}-\alpha)(X_{i2}-\alpha)^3] &= 3v0_{21}v0_{22}, \\
E[(X_{i1}-\alpha)^2(X_{i2}-\alpha)^2] &= 2v0_{21}^2 + v0_{11}v0_{22}, \\
E[(X_{i1}-\alpha)^3(X_{i2}-\alpha)] &= 3v0_{21}v0_{11}.
\end{aligned}
$$

We can work out the form for $U = E[\frac{\partial^2 l^*(\theta)}{\partial\theta^2}]$. The form of the result for U takes pages to show and can be provided upon request. In order to work out the value of $bias(\kappa)$, note

$$
\begin{aligned}
& n \cdot bias(\kappa) = n \cdot bias(\eta^1) = nE[B^{1,k}B^k] - \frac{n}{2}\beta^{1,kl}E[B^kB^l] \\
= \; & E\Big[\frac{\partial f^1(X_i, X_{i+1};\eta_0)}{\partial\eta^k} f^k(X_i, X_{i+1};\eta_0)\Big] + \frac{1}{n}\sum_{i=1}^{n-1}(n-i)E\Big[\frac{\partial f^1(X_{i+1}, X_{i+2};\eta_0)}{\partial\eta^k} f^k(X_1, X_2;\eta_0)\Big] \\
& -\frac{1}{2}E\Big[\frac{\partial^2 f^1(X_1, X_2;\eta_0)}{\partial\eta^k\partial\eta^l}\Big]E[f^k(X_1, X_2;\eta_0)f^l(X_1, X_2;\eta_0)] \\
= \; & E\Big[u_{1i}\frac{\partial g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial\theta^k} \cdot u_{kj}g^j(X_\alpha, X_{\alpha+1};\theta_0)\Big] \\
& -\frac{1}{2}E\Big[u_{1i}\frac{\partial^2 g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial\theta^k\partial\theta^l}\Big]E[u_{ki}g^i(X_\alpha, X_{\alpha+1};\theta_0) \cdot u_{lj}g^j(X_\alpha, X_{\alpha+1};\theta_0)] \\
& +\sum_{\alpha=1}^{n-1}\frac{n-\alpha}{n}E\Big[u_{1i}\frac{\partial g^i(X_\alpha, X_{\alpha+1};\theta_0)}{\partial\theta^k} \cdot u_{kj}g^j(X_1, X_2;\theta_0)\Big].
\end{aligned}
$$

Similar to the one dimension case, if we substitute in our results for each of three parts of our formula and simplify, we will get the results shown in the main part. This process is partially done with the help of using the software Mathematica. Summarization of the key outputs will be provided upon request.

CHAPTER 3. KERNEL SMOOTHED VOLATILITY INDEX ESTIMATION

In this chapter, we first introduce some related works and concepts about volatility in Section 3.1. We also introduce in detail about how to calculate the volatility indices, VXO and VIX, which were introduced by the Chicago Board Option Exchange (CBOE) in 1993 and 2003, respectively. In Section 3.2, we propose our method of estimating the VIX. Essentially, we add a non-parametric smoothing procedure to the selected option prices to improve the CBOE VIX formula. Estimation bias is reduced, as shown in Section 3.3, where we include some simulation results to compare our method and the existing methods.

3.1 Introduction

In this section, we briefly review the development of volatility theories. Some basic concepts and theorems about option pricing and volatility measuring are formally introduced. Formulae for calculating the volatility index adopted by the Chicago Board Option Exchange (CBOE) are introduced.

3.1.1 Stock, Index and Option

With the original debut of stock market that can be traced to as early as 12-th century, companies have used stocks to help raise money. To satisfy the need of different investors, various derivatives like villain options are created. Rational investors are eager to profit by buying cheap stocks and cheap options and then selling them at higher prices, while being exposed to as minimal risk as they can.

One possible measure of market risk is the volatility. In order to compensate for risks investors are exposed to, the stock price is expected to increase through time, or the discounted price is expected to be a martingale in a viable market. One of the earliest stock price models, Black-Scholes, suggested the stock price evolves as a geometric Brownian motion. However, the volatility tends to be mean reverting, namely when the volatility is high, the price tends to revert to its long term mean, while when the volatility is low, the price tends to drift to higher. When market volatility increases, higher rates of return is expected; when the volatility falls, the vice versa.

To better grasp the total market information, many market indexes like the well known Standard and Poor's 500 were introduced. Standard and Poor's 500, or S & P 500, with ticker symbol "SPX", is a stock market index based on a weighted sum with weight proportional to market capitalizations of 500 leading companies publicly traded in the U.S. stock market. It represents the market well and hence can be treated as a good benchmark of market movement.

Similarly, the CBOE introduced volatility indexes to represent the market volatility. These volatility indexes also act as "equities" upon which futures and options are derived. In 1993, the CBOE introduced a volatility index called VXO based on the option prices of the S and P 100, ticker symbol "OEX". VXO is calculated by including eight at-the-money index calls and puts. We sketch it in Section 3.1.3. For more detail, please refer to Whaley (1993).

Bollen and Whaley (2004) showed that the demand to buy out-of-the money and at-the-money SPX puts is a key drive in the movement in SPX implied volatility measures. They took an empirical approach and established an AR(1) type relation between the average implied volatility and the underlying security return, trading volume of the underlying stock, net buying pressure. The net buying pressure for index puts affects the shape of the implied volatility of the S & P 500 most strongly. Meanwhile, the S & P 500 market become more and more liquid than S & P 100. In the early 1990's, the trading column of the S & P 500 option market is about one-fifth of the S & P 100 option market. Ten years later, it is thirteen times as large as the later. Hence, the CBOE moved from VXO to VIX on 09/22/2003 to not only change the fundamental source of option prices from OEX to SPX, but also to include out-of-the-money option prices for estimating the market volatility.

There are varieties of volatilities that can be used to measure the market risk and there are some different methods to estimate each of them. We are mainly interested in improving the estimation for the volatility index VIX. Please refer to Section 3.1.2.

An option in the financial world is a contract that gives its buyer or owner the right, but not the obligation, to buy or to sell a certain amount of financial asset on or before a certain date at a specific price. We call the date the "maturity date" or expiration and the price the "strike price" or "exercise price". The buyer pays a premium to the seller for this right. An option that gives the buyer the right to buy a certain asset at a certain price is called a "call option", and the option that ensure the buyer the right to sell certain asset at a preset price is called a "put option".

The seller or writer of the option has the obligation to fulfill the transaction if the buyer chooses to exercise his or her right, that is to sell certain amount of asset to the buyer at the agreed price if a call option is exercised by the buyer, or to buy certain amount of asset from the buyer at the negotiated price if the put option is exercised by the buyer.

The seller needs to make the following contract specifications:

- ⋆ the type of the option: the buyer has the right to buy (call option) or to sell (put option);
- ⋆ the underlying asset: any financial asset can be used, typically a stock, a bond, or a currency;
- ⋆ the quantity of the underlying asset to be sold or bought, usually 100;
- ⋆ the strike price: the price at which the underlying asset will be transferred between the buyer and solder when the option is exercised;
- ⋆ the maturity date: the date on or before which the buyer of the option can exercise the

right to buy or sell the underlying asset at the strike price; for example, an option that can be only exercised on the maturity date is called a European option, an option that can be exercised at any time before the maturity is called an American option;

- ⋆ the settlement term: whether the actual asset is transferred between the buyer and the seller when the option is exercised or just deliver the equivalent cash amount;

- ⋆ other terms like how the option is quoted based on the market price for the underlying asset into the option price.

Options have been used ever since nineteenth century. However, beginning in 1973, most options were issued with standardized terms and traded by guaranteed clearing houses like the Chicago Board Options Exchange (CBOE). Since then, option tradings has expanded.

Below, we define the option price more rigorously. Consider a European call option with expiration date T, exercise price K, stock price S_t and suppose there's no transaction fee. If the stock price S_T is no greater than the exercise price K, a reasonable buyer of the option would not exercise the option. If the stock price S_T is larger than the exercise price K, the option holder can profit $S_T - K$ by exercising the option. So at time T, the value of the option is

$$max(S_T - K, 0).$$

Intuitively, the fair price of the option at time t can be conjectured to be the current value of the expected value of the option:

$$e^{-r(T-t)} E_t[max(S_T - K, 0)],$$

where r is the risk-free interest rate and E_t is the expectation based on information known at time t. Many models for mimicking the stock price trend have been suggested. Among them is one the most famous Black-Scholes model suggested by Black and Scholes (1973). We will include the pricing of options under the Black-Scholes model below. Please refer to Lamberton and Lapeyre (1996) and Hull (2005) for more related option pricing models and theories.

Now that we have formally defined options with mathematical notations, we move on to discuss option pricing and hedging problems. Some related existing martingale theories are cited for easy reference.

Option valuation within the Black-Scholes framework is based on the concept of perfect replication of contingent claims. By modelling the price of stock (the underlying asset), S_t as a continuous-time diffusion

$$dS_t = S_t(\mu dt + \sigma dB_t)$$

with constant parameters, μ and σ, we will show that an investor can replicate an option's return stream by continuously rebalancing a self-financing portfolio involving stocks and risk-free bonds. By definition, the wealth at time t of the replicating portfolio equals the price of the option. The closed-form expressions for both the option's price and the replicating strategy in the Black-Scholes settings will be derived. Before that, we confirm such a strategy indeed exists by using the famous Girsanov theorem and martingale representation theorem.

3.1.1.1 The Behavior of Prices

In the Black-Scholes model, the behavior of prices is a continuous time model with one risky asset (a share with price S_t at time t) and a riskless asset (with price S_t^0 at time t). We suppose the behavior of S_t^0 to be encapsulated by the following (ordinary) differential equation:

$$dS_t^0 = rS_t^0 dt, \tag{1.1}$$

where r is a non-negative constant. Note that r is an *instantaneous interest rate* and should not be confused with one-period rate in discrete-time models. We set $S_0^0 = 1$, so that $S_t^0 = e^{rt}$ for $t \geq 0$. We assume that the behavior of the stock price is determined by the following stochastic differential equation:

$$dS_t = S_t(\mu dt + \sigma dB_t), \tag{1.2}$$

where μ and σ are two constants and (B_t) is a Standard Brownian motion.

The model is valid on the interval [0,T] where T is the maturity of the option. The information in the market until time s is suggested by $\mathcal{F}_s \triangleq \sigma\{S_t; t \leq s\}$, that means our model is under the natural filtration of a standard Brownian motion.

Remark 3.1.1 *Applying the Itô formula to* $\log S_t$*, it's easy to get a closed-form solution of (1.2):*

$$S_t = S_0 \exp[(\mu - \frac{\sigma^2}{2})t + \sigma B_t]. \tag{1.3}$$

where S_0 *is the spot price observed at time* 0*.*

Remark 3.1.2 *In model (1.2), the basic assumption is the parameter* μ *and* σ *are both constants. Note that constant* σ *implies the market volatility is a constant. It is reasonable to make this assumption for mathematical simplicity when the market is stable. Some similar models used to be considered are* $dS_t = dB_t$ *and* $dS_t = adt + bdB_t$*, where* a*,*b *are both constants. To include more general cases, we sometimes relax the assumptions and consider*

$$\frac{dS_t}{S_t} = \mu(t, S_t) + \sigma(t, S_t)dB_t.$$

According to the definition of Brownian motion, the process S_t in (1.3) has the following properties which express in concrete terms the hypotheses of Black and Scholes on the behavior of the share price :

- continuity of the sample paths;
- independence of the relative increments: if $u \leq t$, S_t/S_u or (equivalently), the relative increment $(S_t - S_u)/S_u$ is independent of the σ-algebra $\sigma(S_v, v \leq u)$;
- stationarity of the relative increments: if $u \leq t$, the law of $(S_t - S_u)/S_u$ is identical to the law of $(S_{t-u} - S_0)/S_0$.

3.1.1.2 Self-financing Strategies

A strategy will be defined as a process $\phi = (\phi_t)_{0 \leq t \leq T} = ((H_t^0, H_t))$ with values in $\mathbb{R}^2$, adapted to the natural filtration $(\mathcal{F}_t)$ of the Brownian motion; the components H_t^0 and H_t are the quantities of riskless asset and risky asset respectively, held in the portfolio at time t. The value of the portfolio at time t is then given by

$$V_t(\phi) = H_t^0 S_t^0 + H_t S_t.$$

The self-financing strategies in the continuous time case is specified by the equality:

$$dV_t(\phi) = H_t^0 dS_t^0 + H_t dS_t.$$

To give a meaning of this equality, we set the condition

$$\int_0^T |H_t^0| dt < \infty \; a.s. \quad \text{and} \quad \int_0^T H_t^2 dt < \infty \; a.s.$$

Then the integral

$$\int_0^T H_t^0 dS_t^0 = \int_0^T H_t^0 re^{rt} dt$$

is well-defined, as is the stochastic integral

$$\int_0^T H_t dS_t = \int_0^T (H_t S_t \mu) dt + \int_0^T \sigma H_t S_t dB_t,$$

since the map $t \to S_t$ is continuous, thus bounded on $[0,T]$ almost surely.

Definition 3.1.1 *A self-financing strategy is defined by a pair ϕ of adapted processes $(H_t^0)_{0 \le t \le T}$ and $(H_t)_{0 \le t \le T}$ satisfying:*

1. $\int_0^T |H_t^0| dt + \int_0^T H_t^2 dt < \infty \quad a.s.$

2. $H_t^0 S_t^0 + H_t S_t = H_0^0 S_0^0 + H_0 S_0 + \int_0^t H_u^0 dS_u^0 + \int_0^t H_u dS_u \; a.s. \quad$ *for all* $\quad t \in [0,T]$.

We denote by $\tilde{S}_t = e^{-rt} S_t$ the discount price of the risky asset.

Proposition 3.1.1 *Let $\phi = ((H_t^0, H_t))_{0 \le t \le T}$ be an adapted process with values in $\mathbb{R}^2$, satisfying $\int_0^T |H_t^0| dt + \int_0^T H_t^2 dt < \infty$ a.s. We set: $V_t(\phi) = H_t^0 S_t^0 + H_t S_t$ and $\tilde{V}_t(\phi) = e^{-rt} V_t(\phi)$. Then, ϕ defines a self-financing strategy if and only if*

$$\tilde{V}_t(\phi) = V_0(\phi) + \int_0^t H_u d\tilde{S}_u \quad \text{a.s.} \tag{1.4}$$

for all $t \in [0,T]$.

Proof : Let us consider a self-financing strategy ϕ. From equality

$$d\tilde{V}_t(\phi) = -r\tilde{V}_t(\phi)dt + e^{-rt}dV_t(\phi)$$

which results from the differentiation of the product of the processes (e^{-rt}) and $(V_t(\phi))$(the cross-variation term $d\langle e^{-r\cdot}, V_\cdot(\phi)\rangle_t$ is null), we deduce

$$\begin{aligned} d\tilde{V}_t(\phi) &= -re^{-rt}(H_t^0 e^{rt} + H_t S_t)dt + e^{-rt}H_t^0 d(e^{rt}) + e^{-rt}H_t dS_t \\ &= H_t(-re^{-rt}S_t dt + e^{-rt}dS_t) \\ &= H_t d\tilde{S}_t, \end{aligned}$$

which yields equality (1.4). The converse is justified similarly. □

We can get inspiration from the general pricing theorem that if the discounted value of the portfolio $(\tilde{V}_t)$ is a martingale, we have the formula $\tilde{V}_t = \mathbf{E}(\tilde{V}_T|\mathcal{F}_t)$, and particularly $C = V_0 = \mathbf{E}(\tilde{V}_T)$. It turns out the discounted price $\tilde{S}_t$ needs to be a martingale. However, (1.2) suggests that S_t is a combination of a martingale and an increasing process. We will resort to Girsanov theorem and consider a probability measure equivalent to the initial probability and under which the discounted price of asset is a martingale. We are then able to design self-financing strategies replicating the option. The following section provides the tools which allow us to apply these methods in continuous time.

3.1.1.3 No Arbitrage and Put/Call Parity

In a liquid financial market, there should be no arbitrage opportunity, i.e. there is no riskless profit available in the market. The notation of arbitrage, the possibility of riskless profit, can be formally defined as

Definition 3.1.2 *An arbitrage strategy is an admissible strategy with zero initial value and non-zero final values.*

Denote the call and put option prices at time t by C_t and P_t respectively. Then when we assume the well accepted assumption of no arbitrage, we have the Put/Call parity formula which was first suggested in Stoll (1969):

$$C_t - P_t = S_t - Ke^{-r(T-t)}.$$

To understand why this parity must hold or there would be arbitrage, let us assume

$$C_t - P_t > S_t - Ke^{-r(T-t)}.$$

If we purchase a share of a stock and a put option, and sell a call option, the current value the asset combination at time t is

$$C_t - P_t - S_t.$$

If this value is positive, we deposit it to the bank to get interest at rate r, or if it is negative, we borrow this amount at rate r. Then, at maturity time T,

- ⋆ If $S_T > K$, the call option would be exercised, which means we deliver the stock and receive cash K; if we withdraw the deposit, the total asset on our hand is now $K + e^{r(T-t)}(C_t - P_t - S_t) > 0$.
- ⋆ If $S_T < K$, we will exercise the put option to get cash amount K; if we empty our bank account, the total asset value is now again $K + e^{r(T-t)}(C_t - P_t - S_t) > 0$.

If $C_t - P_t < S_t - Ke^{-r(T-t)}$, we can buy a stock and a call option, sell a put option, and argue similarly. Therefore, if the parity is broken, we are guaranteed a positive flow of profit at maturity, which violates the no arbitrage assumption. Therefore, the put/call parity must hold.

3.1.1.4 Equivalent Probabilities

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. A probability measure $\mathbf{Q}$ on $(\Omega, \mathcal{A})$ is *absolutely continuous* relative to $\mathbf{P}$ if

$$\forall A \in \mathcal{A} \quad \mathbf{P}(A) = 0 \Rightarrow \mathbf{Q}(A) = 0.$$

Theorem 3.1.1 *$\mathbf{Q}$ is absolutely continuous relative to $\mathbf{P}$ if and only if there exists a non-negative random variable Z on $(\Omega, \mathcal{A})$ such that*

$$\forall A \in \mathcal{A} \quad \mathbf{Q}(A) = \int_A Z(w) d\mathbf{P}(w).$$

Z is called density of $\mathbf{Q}$ relative to $\mathbf{P}$ and sometimes denoted $d\mathbf{Q}/d\mathbf{P}$.

This theorem is a modified version of Radon-Nikodym theorem. The probabilities $\mathbf{P}$ and $\mathbf{Q}$ are *equivalent* if each one is absolutely continuous relative to the other. Note that if $\mathbf{Q}$ is absolutely continuous relative to $\mathbf{P}$, with density Z, then $\mathbf{P}$ and $\mathbf{Q}$ are equivalent if and only if $\mathbf{P}(Z > 0) = 1$. $[\mathbf{Q}(A) \geq \mathbf{Q}(A \cap \{Z \geq \varepsilon\}) \geq \varepsilon \mathbf{P}(A \cap \{Z \geq \varepsilon\})]$

3.1.1.5 The Girsanov Theorem

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbf{P})$ be a probability space equipped with the natural filtration of a standard Browanian motion $(B_t)_{0 \leq t \leq T}$, indexed on the time interval $[0, T]$. The following theorem, which we admit, is known as the Girsanov theorem.

Theorem 3.1.2 (The Girsanov theorem)
Let $(\theta_t)_{0 \leq t \leq T}$ be an adapted process satisfying $\int_0^t \theta_s^2 ds < \infty$ a.s. and such that the process $(L_t)_{0 \leq t \leq T}$ defined by

$$L_t = \exp\Big(- \int_0^t \theta_s dB_s - \frac{1}{2} \int_0^t \theta_s^2 ds \Big)$$

is a martingale. Then, under the probability $\mathbf{P}^{(L)}$ with density L_T relative to $\mathbf{P}$, the process $(W_t)_{0 \leq t \leq T}$ defined by $W_t = B_t + \int_0^t \theta_s ds$, is a standard Brownian motion.

Remark 3.1.3 *A sufficient condition known as Novikov condition for $(L_t)_{0 \leq t \leq T}$ to be a martingale is: $\mathbf{E}(\exp(\frac{1}{2} \int_0^T \theta_t^2 dt)) < \infty$.*

We now make the proof of Girsanov theorem precise. First we state the useful Levy characterization of Brownian motion without proof. A proof can be found in e.g. Karatzas & Shreve (1988), Th 3.3.16.

Theorem 3.1.3 (The Levy Characterization of Brownian Mmotion)
Let $X(t) = \{X_1^{(t)}, \ldots, X_n^{(t)}\}$ be a continuous stochastic process on a probability space $(\Omega, \mathcal{H}, \mathbf{Q})$ with values in $\mathbb{R}^n$. Then the following, $a)$ and $b)$, are equivalent

a) $X(t)$ is a Brownian motion w.r.t. $\mathbf{Q}$, i.e. the law of $X(t)$ w.r.t. $\mathbf{Q}$ is the same as the law of an n-dimensional Brownian motion.

b) (i) $X(t) = \{X_1^{(t)}, \ldots, X_n^{(t)}\}$ is a martingale w.r.t. $\mathbf{Q}$ (and w.r.t. its own filtration) and

(ii) $X_i(t)X_j(t) - \delta_{ij}t$ is a martingale w.r.t $\mathbf{Q}$ (and w.r.t. its own filtration) for all $i, j \in \{1, 2, \ldots, n\}$.

Next we need an auxiliary result about conditional expectation:

Lemma 3.1.1 *Let μ and ν be two probability measures on a measurable space $(\Omega, \mathcal{G})$ such that $d\nu(\omega) = f(\omega)d\mu(\omega)$ for some $f \in L^1(\mu)$. Let X be a random variable on $(\Omega, \mathcal{G})$ such that*

$$\mathbf{E}_\nu[|X|] = \int_\Omega |X(\omega)| f(\omega) d\mu(\omega) < \infty.$$

Let $\mathcal{H}$ be a σ-algebra, $\mathcal{H} \subset \mathcal{G}$. Then

$$\mathbf{E}_\nu[X|\mathcal{H}] \cdot \mathbf{E}_\mu[f|\mathcal{H}] = \mathbf{E}_\mu[fX|\mathcal{H}] \quad a.s. \tag{1.5}$$

Proof: By the definition of conditional expectation we have that if $H \in \mathcal{H}$ then

$$\begin{aligned}\int_H \mathbf{E}_\nu[X|\mathcal{H}] f d\mu &= \int_H \mathbf{E}_\nu[X|\mathcal{H}] d\nu = \int_H X d\nu \\ &= \int_H X f d\mu = \int_H \mathbf{E}_\mu[fX|\mathcal{H}] d\mu\end{aligned} \tag{1.6}$$

On the other hand, we have

$$\begin{aligned}\int_H \mathbf{E}_\nu[X|\mathcal{H}] f d\mu &= \mathbf{E}_\mu[\mathbf{E}_\nu[X|\mathcal{H}] f \cdot \mathbf{1}_H] = \mathbf{E}_\mu[\mathbf{E}_\mu[\mathbf{E}_\nu[X|\mathcal{H}] f \cdot \mathbf{1}_H|\mathcal{H}]] \\ &= \mathbf{E}_\mu[\mathbf{1}_H \mathbf{E}_\nu[X|\mathcal{H}] \cdot \mathbf{E}_\mu[f|\mathcal{H}]] = \int_H \mathbf{E}_\nu[X|\mathcal{H}] \cdot \mathbf{E}_\mu[f|\mathcal{H}] d\mu\end{aligned} \tag{1.7}$$

Combining (1.6) and (1.7) we get

$$\int_H \mathbf{E}_\nu[X|\mathcal{H}] \cdot \mathbf{E}_\mu[f|\mathcal{H}] d\mu = \int_H \mathbf{E}_\mu[fX|\mathcal{H}] d\mu$$

Since this holds for all $H \in \mathcal{H}$, (1.5) follows. $\square$

Proof of Theorem 3.1.2: In view of the Levy characterization of Brownian motion theorem we have to verify that

(i) $W(t) = (W_1(t), \ldots, W_n(t))$ is a martingale w.r.t. $\mathbf{P}^{(L)}$;

(ii) $W_i(t)W_j(t) - \delta_{ij}t$ is a martingale w.r.t $\mathbf{P}^{(L)}$, for all $i, j \in \{1, 2, \ldots, n\}$.

To verify (i) we put $K(t) \triangleq L_t W(t)$ and use Ito's formula to get

$$\begin{aligned} dK_i(t) &= L_t dW_i(t) + W_i(t) dL_t + dW_i(t) dL_t \\ &= L_t (dB_i(t) - W_i(t) \sum_{k=1}^{n} \theta_k(t) dB_k(t)) = L_t \gamma^{(i)}(t) dB(t) \end{aligned}$$

where $\gamma^{(i)}(t) = (\gamma_1^{(i)}(t), \ldots, \gamma_n^{(i)}(t))$, with

$$\gamma_j^{(i)}(t) = \begin{cases} -W_i(t)\theta_j(t) & j \neq i \\ 1 - W_i(t)\theta_i(t) & j = i. \end{cases}$$

Hence $K_i(t)$ is a martingale w.r.t. $\mathbf{P}$. so by the Lemma we get, for $t > s$,

$$\mathbf{E}_{\mathbf{P}^{(L)}}[W_i(t)|\mathcal{F}_s] = \frac{\mathbf{E}[L_t W_i(t)|\mathcal{F}_s]}{\mathbf{E}[L_t|\mathcal{F}_s]} = \frac{\mathbf{E}[K_i(t)|\mathcal{F}_s]}{L_s} = \frac{K_i(s)}{L_s} = W_i(s),$$

which shows that $W_i(t)$ is a martingale w.r.t. $\mathbf{P}^{(L)}$. This proves (i). The proof of (ii) is similar.
$\square$

3.1.1.6 Representation of Brownian Martingales and Option Pricing and Hedging

Let $(B_t)_{0 \leq t \leq T}$ be a standard Brownian motion built on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and let $(\mathcal{F}_t)_{0 \leq t \leq T}$ be its natural filtration. We know that if $(H_t)_{0 \leq t \leq T}$ is an adapted process such that $\mathbf{E}(\int_0^T H_t^2 dt) < \infty$, the process $(\int_0^t H_s dB_s)$ is a square-integrable martingale, null at 0. The following theorem shows that any Brownian martingale can be represented in terms of a s-tochastic integral.

Theorem 3.1.4 (The Martingale Representation Theorem)
Let $(M_t)_{0\leq t\leq T}$ be a square-integrable martingale, with respect to the filtration $(\mathcal{F}_t)_{0\leq t\leq T}$. There exists an adapted process $(H_t)_{0\leq t\leq T}$ such that $\mathbf{E}(\int_0^T H_s^2 ds) < +\infty$ and

$$\forall t \in [0,T] \qquad M_t = M_0 + \int_0^t H_s dB_s \qquad a.s. \tag{1.8}$$

Note that this representation only applies to martingales relative to the *natural* filtration of the Brownian motion. From this theorem, it follows that if U is an $\mathcal{F}_t$-measurable, square-integrable random variable, it can be written as

$$U = \mathbf{E}(U) + \int_0^T H_s dB_s \quad \text{a.s.,}$$

where (H_t) is an adapted process such that $\mathbf{E}(\int_0^T H_s^2 ds) < +\infty$. To prove it, consider the martingale $M_t = \mathbf{E}(U|\mathcal{F})$. To prove the representation theorem, we first establish some auxiliary results.

Lemma 3.1.2 *Fix $T > 0$. The set of random variables*

$$\{\phi(B_{t_1},\ldots,B_{t_n}); t_i \in [0,T], \phi \in C_0^\infty(\mathbb{R}^n), n = 1,2,\ldots\}$$

is dense in $L^2(\mathcal{F}_T,\mathbf{P})$.

Proof : Let $\{t_i\}_{i=1}^\infty$ be a dense subset of $[0,T]$ and for each $n = 1,2,\ldots$ let $\mathcal{H}_n$ be the σ-*algebra* generated by $B_{t_1},\ldots,B_{t_n}$. Then clearly

$$\mathcal{H}_n \subset \mathcal{H}_{n+1}$$

and $\mathcal{F}_T$ is the smallest σ-*algebra* containing all the $\mathcal{H}_n$'s. Choose $g \in L^2(\mathcal{F}_T,\mathbf{P})$. Then by the martingale convergence theorem we have that

$$g = \mathbf{E}[g|\mathcal{F}_T] = \lim_{n\to\infty} \mathbf{E}[g|\mathcal{H}_n]$$

The limit pointwise a.e. ($\mathbf{P}$) and in $L^2(\mathcal{F}_T,\mathbf{P})$. By the Doob-Dynkin Lemma we can write, for each n,

$$\mathbf{E}[g|\mathcal{H}_n] = g_n(B_{t_1},\ldots,B_{t_n})$$

for some Borel measurable function $g_n : \mathbb{R}^n \to \mathbb{R}$. Each such $g_n(B_{t_1},\ldots,B_{t_n})$ can be approximated in $L^2(\mathcal{F}_T,\mathbf{P})$ by functions $\phi_n(B_{t_1},\ldots,B_{t_n})$ where $\phi_n \in C_0^\infty(\mathbb{R}^n)$ and the result follows. $\square$

Lemma 3.1.3 *The linear span of random variables of the type*

$$\exp\{\int_0^T h(t)dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t)dt\}; \qquad h \in L^2[0,T] \tag{1.9}$$

is dense in $L^2(\mathcal{F}_T,\mathbf{P})$.

Proof : Suppose $g \in L^2(\mathcal{F}_T,\mathbf{P})$ is orthogonal (in $L^2(\mathcal{F}_T,\mathbf{P})$) to all functions of the form of (1.9). Then in particular

$$G(\lambda) \triangleq \int_\Omega \exp\{z_1 B_{t_1}(\omega) + \cdots + z_n B_{t_n}\} g(\omega) d\mathbf{P}(\omega) = 0$$

for all $\lambda = (\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^n$ and all $t_1,\ldots,t_n \in [0,T]$. The function $G(\lambda)$ is real analytic in $\lambda \in \mathbb{R}^n$ and hence G has an analytic extension to the complex space $\mathbb{C}^n$ given by

$$G(z) = \int_\Omega \exp\{z_1 B_{t_1}(\omega) + \cdots + z_n B_{t_n}\} g(\omega) d\mathbf{P}(\omega)$$

for all $z = (z_1,\ldots,z_n) \in \mathbb{C}^n$. Since $G = 0$ on $\mathbb{R}^n$ and G is analytic, $G = 0$ on $\mathbb{C}^n$. In particular, $G(iy_1, iy_2,\ldots,iy_n) = 0$ for all $y = (y_1,\ldots,y_n) \in \mathbb{R}^n$. But then we get, for $\phi \in C_0^\infty(\mathbb{R}^n)$,

$$\begin{aligned}
& \int_\Omega \phi(B_{t_1},\ldots,B_{t_n}) g(\omega) d\mathbf{P}(\omega) \\
= \; & \int_\Omega (2\pi)^{-n/2} (\int_{\mathbb{R}_n} \hat{\phi}(y) e^{i(y_1 B_{t_1} + \cdots + y_n B_{t_n})} dy) g(\omega) d\mathbf{P}(\omega) \\
= \; & (2\pi)^{-n/2} \int_{\mathbb{R}_n} \hat{\phi}(y) (\int_\Omega e^{i(y_1 B_{t_1} + \cdots + y_n B_{t_n})} g(\omega) d\mathbf{P}(\omega)) dy \\
= \; & (2\pi)^{-n/2} \int_{\mathbb{R}_n} \hat{\phi}(y) G(iy) dy = 0,
\end{aligned} \tag{1.10}$$

where

$$\hat{\phi}(y) = (2\pi)^{-n/2} \int_{\mathbb{R}_n} \phi(x) e^{-i(x,y)} dx$$

is the Fourier transform of ϕ and we have used the inverse Fourier transform theorem

$$\phi(x) = (2\pi)^{-n/2} \int_{\mathbb{R}_n} \hat{\phi}(x) e^{-i(x,y)} dy.$$

By (1.10) and the last Lemma g is orthogonal to a dense subset of $L^2(\mathcal{F}_T, \mathbf{P})$ and we conclude that $g = 0$. Therefore the linear span of the functions in (1.9) must be dense in $L^2(\mathcal{F}_T, \mathbf{P})$ as claimed. □

Theorem 3.1.5 (The Ito Representation Theorem)

Let $F \in L^2(\mathcal{F}_T, \mathbf{P})$. Then there exists a unique adapted process $f(t, \omega)$ such that $\mathbf{E}(\int_0^T f_s^2 ds) < +\infty$ and

$$F(\omega) = \mathbf{E}[F] + \int_0^T f(t, \omega) dB_t(\omega). \tag{1.11}$$

Proof : First assume that F has the form (1.9), i.e.

$$F(\omega) = \exp\{\int_0^T h(t) dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t) dt\}$$

for some $h(t) \in L^2[0, T]$.

Define

$$Y_t(\omega) = \exp\{\int_0^t h(s) dB_s(\omega) - \frac{1}{2}\int_0^t h^2(s) ds\} \qquad t \in [0, T].$$

Then by Ito's formula

$$dY_t = Y_t(h(t)dB_t - \frac{1}{2}h^2(t)dt) + \frac{1}{2}Y_t(h(t)dB_t)^2 = Y_t h(t) dB_t$$

so that

$$Y_t = 1 + \int_0^t Y_s h(s) dB_s; \qquad t \in [0, T].$$

Therefore

$$F = Y_T = 1 + \int_0^T Y_s h(s) dB_s$$

and hence $\mathbf{E}[F] = 1$. So (1.11) holds in this case. By linearity (1.11) also holds for linear combinations of function of the form (1.9). So if $F \in L^2(\mathcal{F}_T, \mathbf{P})$ is arbitrary, we approximate

F in $L^2(\mathcal{F}_T, \mathbf{P})$ by linear combinations F_n of functions of the form (1.9). Then for each n we have

$$F_n(\omega) = \mathbf{E}[F_n] + \int_0^T f_n(s,\omega)dB_s(\omega),$$

where f_n is an adapted process such that $\mathbf{E}(\int_0^T f_n^2 ds) < +\infty$. By the Ito isometry

$$\begin{aligned} \mathbf{E}[(F_n - F_m)^2] &= \mathbf{E}[(\mathbf{E}[F_n - F_m] + \int_0^T (f_n - f_m)dB)^2] \\ = \quad & (\mathbf{E}[F_n - F_m])^2 + \int_0^T \mathbf{E}[(f_n - f_m)^2]dt \to 0 \qquad \text{as} \quad n,m \to \infty \end{aligned}$$

so $\{f_n\}$ is a Cauchy sequence in $L^2([0,T] \times \Omega)$ and hence converges to some $f \in L^2([0,T] \times \Omega)$. Since f_n adapted and $\mathbf{E}(\int_0^T f_n^2 ds) < +\infty$, we have f adapted and $\mathbf{E}(\int_0^T f^2 ds) < +\infty$, again using the Ito isometry we see that

$$F = \lim_{n\to\infty} F_n = \lim_{n\to\infty} (\mathbf{E}[F_n] + \int_0^T f_n dB) = E[F] + \int_0^T f dB,$$

the limit being taken in $L^2(\mathcal{F}_T, \mathbf{P})$. Hence the representation (1.11) holds for all $F \in L^2(\mathcal{F}_T, \mathbf{P})$. The uniqueness follows from the Ito isometry: Suppose

$$F(\omega) = \mathbf{E}[F] + \int_0^T f_1(t,\omega)dB_t(\omega) = \mathbf{E}[F] + \int_0^T f_2(t,\omega)dB_t(\omega)$$

with f_1, f_2 are adapted processes and satisfy $\mathbf{E}(\int_0^T f_i^2(s)ds) < +\infty \quad i = 1,2$. Then

$$0 = \mathbf{E}[(\int_0^T (f_1(t,\omega) - f_2(t,\omega))dB_t(\omega))^2] = \int_0^T \mathbf{E}[(f_1(t,\omega) - f_2(t,\omega))^2]dt$$

and therefore $f_1(t,\omega) = f_2(t,\omega)$ for a.a. $(t,\omega) \in [0,T] \times \Omega$. $\square$

Proof of Theorem 3.1.4: By the Ito representation theorem applied to $t = T, F = M_t$, we have that for all t there exists a unique $h^{(t)}(s,\omega) \in L^2(\mathcal{F}_t, \mathbf{P})$ such that

$$M_t(\omega) = \mathbf{E}[M_t] + \int_0^t h^{(t)}(s,\omega)dB_s(\omega) = \mathbf{E}[M_0] + \int_0^t h^{(t)}(s,\omega)dB_s(\omega)$$

Now assume $0 \leq t_1 < t_2$. Then

$$M_{t_1} = \mathbf{E}[M_{t_2}|\mathcal{F}_{t_1}] = \mathbf{E}[M_0] + \mathbf{E}[\int_0^{t_2} h^{(t_2)}(s,\omega)dB_s(\omega)|\mathcal{F}_{t_1}] = \mathbf{E}[M_0] + \int_0^{t_1} h^{(t_2)}(s,\omega)dB_s(\omega)$$

But we also have

$$M_{t_1} = \mathbf{E}[M_0] + \int_0^{t_1} h^{(t_1)}(s,\omega)dB_s(\omega)$$

Hence, we get that

$$0 = \mathbf{E}[(\int_0^{t_1}(h^{(t_1)} - h^{(t_2)})dB)^2] = \int_0^{t_1}\mathbf{E}[(h^{(t_1)} - h^{(t_2)})^2]ds$$

and therefor

$$h^{t_1}(s,\omega) = h^{t_2}(s,\omega) \qquad \text{for a.a. } (s,\omega) \in [0,t_1]\times\Omega.$$

So we can define $H(s,\omega)$ for a.a. $s \in [0,\infty)\times\Omega$ by setting

$$H(s,\omega) = h^{(N)}(s,\omega) \qquad \text{if } s \in [0,N]$$

and then we get

$$M_t = \mathbf{E}[M_0] + \int_0^t h^{(t)}(s,\omega)dB_s(\omega) = M_0 + \int_0^t H(s,\omega)dB_s(\omega) \quad \text{for all } t \geq 0.$$

□

3.1.1.7 A Probability under which $(\tilde{S}_t)$ is a Martingale

Now we will consider the model introduced in the Section 2. We will prove that there exists a probability equivalent to **P**, under which the discounted share price $\tilde{S}_t = e^{-rt}S_t$ is a martingale. From the stochastic differential equation satisfied by (S_t), we have

$$\begin{aligned} d\tilde{S}_t &= -re^{-rt}S_t dt + e^{-rt}dS_t \\ &= \tilde{S}_t((\mu - r)dt + \sigma dB_t). \end{aligned}$$

Consequently, if we set $W_t = B_t + (\mu - r)t/\sigma$, then we have

$$d\tilde{S}_t = \tilde{S}_t \sigma dW_t. \tag{1.12}$$

From Girsanov theorem, with $\theta_t = (\mu - r)/\sigma$, there exists a probability $\mathbf{P}^*$ equivalent to **P** under which $(W_t)_{0\leq t\leq T}$ is a standard Brownian motion. We will admit that the definition of the stochastic integral is invariant by change of equivalent probability. Then under the probability $\mathbf{P}^*$, we deduce from equality (1.12) that $(\tilde{S}_t)$ is a martingale and that

$$\tilde{S}_t = \tilde{S}_0 \exp(\sigma W_t - \sigma^2 t/2).$$

3.1.1.8 Pricing

In this section, we will focus on European options. A European option will be defined by a non-negative, $\mathcal{F}_t$-measurable, random variable h. Quite often, h can be written as $f(S_T)$, ($f(x) = (x-K)_+$ in the case of a call, $f(x) = (K-x)_+$ in the case of a put). We define the option value by a replication argument, and limit our study to the following admissible strategies:

Definition 3.1.3 *A strategy $\phi = (H_t^0, H_t)_{0 \leq t \leq T}$ is admissible if it is self-financing and if the discounted value $\tilde{V}_t(\phi) = H_t^0 + H_t \tilde{S}_t$ of the corresponding portfolio is, for all t, non-negative and such that $\sup_{t\in[0,T]} \tilde{V}_t$ is square-integrable under $\mathbf{P}^*$.*

An option is said to be replicable if its payoff at maturity is equal to the final value of an admissible strategy. It is clear that, for the option defined by h to be replicable, it is necessary that h should be square-integrable under $\mathbf{P}^*$. In the case of a call $(h = (S_T - K)_+)$, this property indeed holds since $E^*(S_T^2) < \infty$; note that in the case of a put, h is even bounded.

Theorem 3.1.6 *In the Black-Scholes model, any option defined by a non-negative, $\mathcal{F}_T$-measurable random variable h, which is square-integrable under the probability $\mathbf{P}^*$, is replicable and the value at time t of any replicating portfolio is given by*

$$V_t = \mathbf{E}^*(e^{-r(T-t)} h | \mathcal{F}_t).$$

Thus, the option value at time t can be naturally defined by the expression $\mathbf{E}^*(e^{-r(T-t)} h | \mathcal{F}_t)$.

Proof : First, assume that h is replicable, i.e. there is an admissible strategy (H^0, H), replicating the option. The value at time t of the portfolio (H_t^0, H_t) is given by

$$V_t = H_t^0 S_t^0 + H_t S_t,$$

and, by hypothesis, we have $V_T = h$. Let $\tilde{V}_t = V_t e^{-rt}$ be the discounted value

$$\tilde{V}_t = H_t^0 + H_t \tilde{S}_t.$$

Since the strategy is self-financing, we get from Proposition 2.1 and equality (1.12)

$$\begin{aligned}\tilde{V}_t &= V_0 + \int_0^t H_u d\tilde{S}_u \\ &= V_0 + \int_0^t H_u \sigma \tilde{S}_u dW_u.\end{aligned}$$

Under the probability $\mathbf{P}^*$, $\sup_{t\in[0,T]} \tilde{V}_t$ is square-integrable, by definition of admissible strategies. Furthermore, the preceding equality shows that the process $(\tilde{V}_t)$ is a stochastic integral relative to (W_t). It follows that $(\tilde{V}_t)$ is a square-integrable martingale under $\mathbf{P}^*$. Hence

$$\tilde{V}_t = \mathbf{E}^*(\tilde{V}_T|\mathcal{F}_t),$$

and consequently

$$V_t = \mathbf{E}^*(e^{-r(T-t)}h|\mathcal{F}_t). \tag{1.13}$$

So we have proved that if a portfolio (H^0, H) replicates that the option defined by h, its value is given by equality (1.13). To complete the proof the theorem, it remains to show that the option is indeed replicable, i.e. to find some processes (H_t^0) and (H_t) defining an admissible strategy, such that

$$H_t^0 S_t^0 + H_t S_t = \mathbf{E}^*(e^{-r(T-t)}h|\mathcal{F}_t).$$

Under the probability $\mathbf{P}^*$, the process defined $M_t = \mathbf{E}^*(e^{-rT}h|\mathcal{F}_t)$ is a square-integrable martingale. The filtration $(\mathcal{F}_t)$, which is the natural filtration of (B_t), is also the natural filtration of (W_t) and, from the theorem of representation of Brownian martingales, there exists an adapted process $(K_t)_{0\leq t\leq T}$ such that $\mathbf{E}^*(\int_0^T K_s^2 ds) < \infty$ and

$$\forall t \in [0,T] \qquad M_t = M_0 + \int_0^t K_s dW_s \quad a.s.$$

The strategy $\phi = (H^0, H)$, with $H_t = K_t/(\sigma\tilde{S}_t)$ and $H_t^0 = M_t - H_t\tilde{S}_t$, is then, from Proposition 2.1 and equality (1.12), a self-financing strategy; its value at time t is given by

$$V_t(\phi) = e^{rt}M_t = \mathbf{E}^*(e^{-r(T-t)}h|\mathcal{F}_t).$$

This expression clearly shows that $V_t(\phi)$ is a non-negative random variable, with $\sup_{0\leq t\leq T} V_t(\phi)$ square-integrable under $\mathbf{P}^*$ and that $V_T(\phi) = h$. So we have found an admissible strategy replicating h. □

Remark 3.1.4 *When the random variable h can be written as $h = f(S_T)$, we can express the option value V_t at time t as a function of t and S_t. We have indeed*

$$\begin{aligned} V_t &= \mathbf{E}^*(e^{-r(T-t)} f(S_T)|\mathcal{F}_t) \\ &= \mathbf{E}^*(e^{-r(T-t)} f(S_t e^{r(T-t)} e^{\sigma(W_T - W_t) - (\sigma^2/2)(T-t)})|\mathcal{F}_t). \end{aligned}$$

The random variable S_t is $\mathcal{F}_t$-measurable and, under $\mathbf{P}^*$, $W_T - W_t$ is independent of $\mathcal{F}_t$. Therefore, we can deduce that

$$V_t = F(t, S_t),$$

where

$$F(t,x) = \mathbf{E}^*(e^{-r(T-t)} f(xe^{r(T-t)} e^{\sigma(W_T - W_t) - (\sigma^2/2)(T-t)})). \tag{1.14}$$

Since, under $\mathbf{P}^*$, $W_T - W_t$ is a zero-mean normal variable with variance $T - t$

$$F(t,x) = e^{-r(T-t)} \int_{-\infty}^{+\infty} f(xe^{(r-\sigma^2/2)(T-t)+\sigma y\sqrt{T-t}}) \frac{e^{-y^2/2} dy}{\sqrt{2\pi}}.$$

F can be calculated explicitly for calls and puts. If we choose the case of the call, where $f(x) = (x - K)_+$, we have, from (1.14)

$$\begin{aligned} F(t,x) &= \mathbf{E}^*(e^{-r(T-t)}(xe^{(r-\sigma^2/2)(T-t)+\sigma(W_T - W_t)} - K)_+) \\ &= \mathbf{E}^*(xe^{\sigma\sqrt{\theta}g - \sigma^2\theta/2} - Ke^{-r\theta})_+ \\ &= \mathbf{E}(xe^{\sigma\sqrt{\theta}g - \sigma^2\theta/2} - Ke^{-r\theta})_+ \end{aligned}$$

where g is a standard Gaussian variable and $\theta = T - t$.

Let us set

$$d_1 = \frac{\log(x/K) + (r + \sigma^2/2)\theta}{\sigma\sqrt{\theta}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{\theta}.$$

Using these notations, we have

$$\begin{aligned} F(t,x) &= \mathbf{E}[(xe^{\sigma\sqrt{\theta}g-\sigma^2\theta/2} - Ke^{-r\theta})\mathbf{1}_{\{g+d_2\geq 0\}}] \\ &= \int_{-d_2}^{+\infty}(xe^{\sigma\sqrt{\theta}y-\sigma^2\theta/2} - Ke^{-r\theta})\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy \\ &= \int_{-\infty}^{d_2}(xe^{-\sigma\sqrt{\theta}y-\sigma^2\theta/2} - Ke^{-r\theta})\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy. \end{aligned}$$

Writing this expression as the difference of two integrals and in the first one using the change of variable $z = y + \sigma\sqrt{\theta}$, *we obtain*

$$F(t,x) = xN(d_1) - Ke^{-r\theta}N(d_2), \tag{1.15}$$

where

$$N(d) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d} e^{-x^2/2}dx.$$

Using identical notations and through similar calculations, the price of the put is equal to

$$F(t,x) = Ke^{-r\theta}N(-d_2) - xN(-d_1). \tag{1.16}$$

Remark 3.1.5 *One of the main features of the Black-Scholes model is the fact that the pricing formulae, as well as the hedging formulae we will give later, depend on only one non-observable parameter:* σ, *called 'volatility' by practitioners (the drift parameter* μ *disappears by change of probability). In practice, two methods are used to evaluate* σ*:*

1. *The historical method: in the present model:* $\sigma^2 T$ *is the variance of* $\log(S_T)$ *and the variables* $\log(S_T/S_0)$, $\log(S_{2T}/S_T)$, $\cdots$, $\log(S_{NT}/S_{(N-1)T})$ *are independent and identically distributed. Therefore,* σ *can be estimated by statistical means using the asset prices observed in the past (for example by calculating empirical variances; cf. Dacunha-Castelle and Duflo (1986), Chapter 5).*

2. *The 'implied' method: some options are quoted on organized markets; the price of options (calls and puts) being an increasing function of* σ, *we can associate an 'implied' volatility to each quoted option, by inversion of the Black-Scholes formula. Once the model is identified, it can be used to elaborate hedging schemes.*

In those problems concerning volatility, one is soon confronted with the imperfections of the Black-Scholes model. Important differences between historical volatility and implied volatility are observed, the later seeming to depend upon the strike price and the maturity. In spite of these incoherences, the model is considered as a reference by practitioners.

3.1.1.9 Hedging Calls and Puts

In the proof of Theorem 4.1, we referred to the theorem of representation of Brownian martingales to show the existence of a replicating portfolio. In practice, a theorem of existence is not satisfactory and it is essential to be able to build a real replicating portfolio to hedge an option.

When the option is defined by a random variable $h = f(S_T)$, we shall show that it is possible to find an explicit hedging portfolio. A replicating portfolio must have, at any time t, a discounted value equal to

$$\tilde{\mathrm{V}}_t = e^{-rt}F(t,S_t),$$

where F is the function defined by equality (1.14). Under large hypothesis on f (and, in particular, in the case of calls and puts where we have the closed-form solutions of Remark 4.1), we see that the function F is of class C^∞ on $[0,T]\times \mathbb{R}$. If we set

$$\tilde{\mathrm{F}}(t,x) = e^{-rt}F(t,xe^{rt}),$$

we have $\tilde{\mathrm{V}}_t = \tilde{\mathrm{F}}(t,\tilde{\mathrm{S}}_t)$ and, for $t < T$, from the Itô formula

$$\begin{aligned}\tilde{\mathrm{F}}(t,\tilde{\mathrm{S}}_t) &= \tilde{\mathrm{F}}(0,\tilde{\mathrm{S}}_0) + \int_0^t \frac{\partial \tilde{\mathrm{F}}}{\partial x}(u,\tilde{\mathrm{S}}_u)d\tilde{\mathrm{S}}_u \\ &\quad + \int_0^t \frac{\partial \tilde{\mathrm{F}}}{\partial t}(u,\tilde{\mathrm{S}}_u)du + \int_0^t \frac{1}{2}\frac{\partial^2 \tilde{\mathrm{F}}}{\partial x^2}(u,\tilde{\mathrm{S}}_u)d\langle \tilde{\mathrm{S}},\tilde{\mathrm{S}}\rangle_u\end{aligned}$$

From equality $d\tilde{\mathrm{S}}_t = \sigma\tilde{\mathrm{S}}_t dW_t$, we deduce

$$d\langle \tilde{\mathrm{S}},\tilde{\mathrm{S}}\rangle_u = \sigma^2\tilde{\mathrm{S}}_u^2 du,$$

so that $\tilde{\mathrm{F}}(t,\tilde{\mathrm{S}}_t)$ can be written as

$$\tilde{\mathrm{F}}(t,\tilde{\mathrm{S}}_t) = \tilde{\mathrm{F}}(0,\tilde{\mathrm{S}}_0) + \int_0^t \sigma\frac{\partial \tilde{\mathrm{F}}}{\partial x}(u,\tilde{\mathrm{S}}_u)\tilde{\mathrm{S}}_u dW_u + \int_0^t K_u du.$$

Since $\tilde{F}(t,\tilde{S}_t)$ is a martingale under $\mathbf{P}^*$, the process K_u is necessarily null. Hence

$$\begin{aligned}\tilde{F}(t,\tilde{S}_t) &= \tilde{F}(0,\tilde{S}_0)+\int_0^t \sigma\frac{\partial\tilde{F}}{\partial x}(u,\tilde{S}_u)\tilde{S}_u dW_u \\ &= \tilde{F}(0,\tilde{S}_0)+\int_0^t \frac{\partial\tilde{F}}{\partial x}(u,\tilde{S}_u)d\tilde{S}_u.\end{aligned}$$

The natural candidate for the hedging process H_t is then

$$H_t = \frac{\partial\tilde{F}}{\partial x}(t,\tilde{S}_t) = \frac{\partial F}{\partial x}(t,S_t).$$

If we set $H_t^0 = \tilde{F}(t,\tilde{S}_t) - H_t\tilde{S}_t$, the portfolio (H_t^0, H_t) is self-financing and its discounted value is indeed $\tilde{V}_t = \tilde{F}(t,S_t)$.

Remark 3.1.6 *The preceding argument shows that it is not absolutely necessary to use the theorem of representation of Brownian martingales to deal with options of the form $f(S_T)$.*

Remark 3.1.7 *In the case of a call, we have, using the same notations as in Remark 4.1, by equality (1.15), we have*

$$\frac{\partial F}{\partial x}(t,x) = N(d_1),$$

and in the case of a put, by equality (1.16), we have

$$\frac{\partial F}{\partial x}(t,x) = -N(-d_1).$$

This quantity is often called the 'delta' of the option by practitioners. More generally, when the value at time t of a portfolio can be expressed as $\Psi(t,S_t)$, the quantity $(\partial\Psi/\partial x)(t,S_t)$, which measures the sensitivity of the portfolio with respect to the variations of the asset price at time$t$, is called the'delta' of the portfolio, 'gamma' refers to the second-order derivative $(\partial^2\Psi/\partial x^2)(t,S_t)$,'theta' to the derivative with respect to time and 'vega' to the derivative of Ψ with respect to the volatility σ.

3.1.2 Various Volatility Measures

Below we introduce some frequently used volatility concepts.

Historical Volatility refers to the standard deviation of asset returns in percentage term based on observed historical asset prices over a certain period. Let $\{S_i\}_{0\le i\le n}$ be the observed asset prices between time t_1 and t_2. Then the asset returns are conventionally defined as

$$r_i = \frac{S_i - S_{i-1}}{S_{i-1}}, \text{ or } r_i = \log\{\frac{S_i}{S_{i-1}}\}.$$

We define the historical volatility to be

$$HV = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2}, \tag{1.17}$$

where $\bar{r} = \frac{1}{n-1}\sum_{i=1}^{n} r_i$ is the average return. Here, the historical return HV can be daily return, weekly return, monthly return, or for any desired repeating periods. And the annualized historical volatility would be[1]

$$AHV = HV\sqrt{\frac{252}{\text{number of trading days within sample period}}}.$$

Note, historical variance rather than the historical volatility is assumed to be addable, so we get the historical variance annualized first, and then take the square root of it to get the above annualized historical volatility.

Implied Volatility is the volatility of the underlying asset derived through its option price. Under that derived volatility value, the same current option price is supposed to be realized. There are both model based and model free ways to derive it.

If we have a stochastic differential model for the asset price in mind, there will be a formula linking the option price C and the asset volatility σ,

$$C = f(\sigma, S_t, K, T, r).$$

Note, the option price depend on but may not limited to current stock price S, exercise price K, time to expire T, and interest rate r. Merton (1973) proved that the option price is a monotonically increasing function of the volatility. Hence, there exists a one to one mapping between

[1] Adopt the typical average trading days/weeks: 21 days for a month, 252 days for a year, and 52 weeks for a year.

sigma and C. Denote IV to be the unique value corresponds to one specific value of C, and it is called the implied volatility under the option price C.

Model free implied volatility evolved along with the theories about derivative of volatility. Interested readers can refer to papers mentioned in the section 4.2.1 for related existing works.

Integrated Volatility is essentially the quadratic variation of the log asset price process. Assume the risk neutral dynamic of the asset price follows the following stochastic differential equation:

$$dS_t/S_t = \mu dt + \sigma_t dW_t,$$

where $\mu \in \mathbb{R}^+$ is the expected instantaneous asset return, $\{\sigma_t\}$ is the instantaneous volatility of the asset, which can be a constant, a deterministic function of time t, or even a stochastic process like in Heston stochastic volatility model. Then the integrated volatility is defined as

$$V = \frac{1}{T}\int_0^T \sigma_t^2 dt.$$

When there is jump, say the process satisfies

$$dS_t/S_t = \mu dt + \sigma_t dW_t + (J_t - 1)dN_t,$$

where $\{N_t\}$ is Poisson process with parameter λ and $\{J_t\}$ are independent and identically distributed positive valued random distributions, representing jump intensity and jump size respectively. Then the integrated volatility is defined as

$$V = V_c + V_d = \frac{1}{T}\int_0^T \sigma_t^2 dt + \frac{1}{T}\sum_{j=1}^{N_T} (\log J_t)^2.$$

Please refer to Broadie and Jain (2008) and Joshi (2008) for details.

We briefly discuss these volatilities under the Black-Scholes model and the Merton Model with jump in the following subsections. These two processes will be used in our simulation.

3.1.2.1 Black-Scholes Formula

One of the most simple and popular model for option price is the Black-Scholes model. In a viable market, suppose the stock price follows [2]

$$dS_t/S_t = rdt + \sigma dW_t,$$

where $r, \sigma \in \mathbb{R}^+$, r is the risk free interest, and $\{W_t\}$ is the standard Brownian motion. This process admits a closed form solution

$$\log S_T = \log S_0 + rT - \frac{1}{2}\sigma^2 T + \sigma B_T.$$

Under this model, at each i, the log returns

$$\log \frac{S_{(i+1)T}}{S_{iT}} \sim N(rT - \frac{1}{2}\sigma^2 T, \sigma^2 T),$$

and they are independently and identically distributed. Hence, we can use historical means of the log price ratio to estimate the volatility σ by

$$\sqrt{\frac{2}{T}\left(rT - \frac{1}{n}\sum_{i=1}^{n}\log\frac{S_{(i+1)T}}{S_{iT}}\right)}.$$

Alternatively, we can use the historical volatility formula 1.17 to estimate σ.

Let T be the maturity time of a call option and K be the exercise price. The corresponding fair price for this call option at time zero is

$$C_{BS} = BS(S_0, \sigma, r, T, K) = E[e^{-rT}(S_T - K)_+ | \mathcal{F}_0] = S_0 N(d_1) - Ke^{-rT} N(d_2),$$

where

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \text{ and } d_2 = d_1 - \sigma\sqrt{T}.$$

With this known closed form for the option price, we can get the inverse function of it when related to the volatility:

$$\hat{\sigma} = C_{BS}^{-1}(C_{BS}; S_0, r, T, K).$$

[2]The drift parameter for the original Black-Scholes model can be any $\mu \in \mathbb{R}^+$. Here we adopt it to equal the risk free interest for clear presentation. In fact, the more general one can be transformed to this based on Girsonov theorem, and μ will not appear in the option price formula.

We note this inverse function exists because, as mentioned in the implied volatility section, Merton (1973) proved that the option price is a monotonically increasing function of the volatility. We call the $\hat{\sigma}$ calculated by this formula with known option price, initial stock price, etc. as the implied volatility.

As for the integrated volatility, because the diffusion parameter here is a constant σ, it is simply

$$V = \frac{1}{T}\int_0^T \sigma^2 dt = \sigma^2,$$

3.1.2.2 Merton Model with Jumps

Merton (1976) first introduced the so called Merton jump model. Assume the dynamic of stock price can be described by [3]

$$dS_t/S_t^- = (r - \lambda(m-1))dt + \sigma dW_t + (J_t - 1)dN_t,$$

where $r, \lambda, m \in \mathbb{R}^+$, r is the risk free interest rate, σ^2 is the instantaneous variance of the return, $\{W_t\}$ is the standard Brownian motion, $\{N_t\}$ is a Poisson process with parameter λ, and $\{\log(J_t)\} \overset{i.i.d.}{\sim} N(\log(m+1) - \frac{1}{2}v^2, v^2)$. And the Brownian motion $\{W_t\}$, jump size $\{J_t\}$ and jump frequency $\{N_t\}$ are all independent of each other.

Note that when a jump occurs, S_t^- change by $S_t^-(J_t - 1)$, or change to $S_t^- J_t$. By the Itô formula,

$$d\log S_t = (r - \lambda(m-1))dt + \sigma dW_t + \log J_t dN_t.$$

Please refer to Joshi (2008). For each t, let Z_t be a standard normal random variable, and they are independent. Note that $J_t \overset{d}{\sim} me^{\{-\frac{1}{2}v^2 + vZ_t\}}$, we have

$$\begin{aligned}
\log S_T &= \log S_0 + (r - \lambda(m-1))T + \sigma W_T + \sum_{j=1}^{N_T} \log J_j \\
&= \log S_0 + (r - \lambda(m-1))T + \sigma\sqrt{T}Z_0 + N_T \log m - \frac{1}{2}v^2 N_T + v\sum_{j=1}^{N_T} Z_j \\
&= \log\left(S_0 m^{N_t} e^{-\lambda(m-1)T}\right) + rT - \frac{1}{2}(\sigma^2 T + v^2 N_T) + \sqrt{\sigma^2 T + v^2 N_T}Z^*,
\end{aligned}$$

[3] The drift parameter is adopted to ensure $\{e^{-rt}S_t\}$ is a martingale.

where Z^* is the standard normal distribution. Given N_T, this corresponds to the stock price in a Black-Scholes model with initial price $S_0 m^{N_T} e^{-\lambda(m-1)T}$ and volatility $\sigma^2 + v^2 N_T/T$. Plugging into the Black-Scholes option price formula and then take expectation with respect to N_T, we arrive at the closed form option price formula:

$$F(S,\tau) = \sum_{n=0}^{\infty} \frac{exp^{-\lambda_0 T}(\lambda_0 T)^n}{n!} BS(S_0, \sigma_n, r_n, T, K),$$

where $\lambda_0 = \lambda m$, $r_n = r - \lambda(m-1) + \frac{n}{T} log(m)$, $\sigma_n = \sqrt{\sigma^2 + \frac{n}{T} v^2}$.

Under the Merton jump model, the fair price for the integrated volatility, or the expected value of it, is

$$\begin{aligned} E(V) &= E[\frac{1}{T}\int_0^T \sigma^2 ds + \frac{1}{T}\sum_{j=1}^{N_T}(\log J_j)^2] \\ &= \sigma^2 + \lambda\left((\log m - \frac{1}{2}v^2)^2 + v^2\right). \end{aligned}$$

3.1.3 VXO - The Early Volatility Index

The Chicago Board Option Exchange (CBOE) Market Volatility index, ticker "VXO", was introduced by Whaley (1993) based on option (American style) prices of Standard & Poor's 100, or S & P 100 (ticker symbol "OEX"). It is defined as an interpolate of the implied volatility of eight near-the-money OEX options, four options each from the nearby and the second nearby OEX option series. It represents a 30 calendar day implied volatility. We enumerate some related definitions below before we present the VXO formula.

- ⋆ Near-the-money options: an option with strike price close to the current market price of the corresponding underlying asset.

- ⋆ Nearby OEX option series: the options with different strike prices which have the shortest time to expiration but with at least eight calendar days to expiration.

- ⋆ Second nearby series: the options with different strike prices which expires later but most adjacent to the Nearby OEX option series.

- ⋆ Let N_c be the number of calender days. Then, the number of trading days N_t is calculated as

$$N_t = N_c - 2 * int(N_c/7),$$

where the second term on the right represents the total number of weekend days.

- ⋆ The implied volatility based on each option is calculated by the Cox-Ross-Rubinstein (1979) binomial tree method with exact cash dividend series. Please see the simulation study in Harvey and Whaley (1992) to find out in detail how to calculate the European-style Black-Scholes implied volatility with dividend adjustment and the approximate American-style option applied volatility with constant dividend yield rate. The interest rate used in the calculation is the 30-day treasury yield.

- ⋆ Assume the total volatility over the option's remaining life is the same, we naturally have the relation:

$$\sigma_t = \sigma_c \sqrt{\frac{N_c}{N_t}},$$

where $\sigma_c(\sigma_t)$ is the calendar-day (trading-day) implied volatility rate.

- ⋆ Denote X_l and X_u to be the exercise price just below and above the current index level, S. Here, S is adopted as the average of the bid/ask price. Use the following symbols, with superscript to denote exercise price and subscript c and p to show option type (call or put) and expiration date (nearby or second nearby), to indicate the eight implied volatilities:

	First Nearby (1)		Second Nearby (2)	
	Call	Put	Call	Put
$X_l(< S)$	$\sigma_{c,1}^{X_l}$	$\sigma_{p,1}^{X_l}$	$\sigma_{c,2}^{X_l}$	$\sigma_{p,2}^{X_l}$
$X_u(\geq S)$	$\sigma_{c,1}^{X_u}$	$\sigma_{p,1}^{X_u}$	$\sigma_{c,2}^{X_u}$	$\sigma_{p,2}^{X_u}$

Then, because the biases of the calculated put/call implied volatilities caused by possible time-delay of reported market price are equal[4], we define:

$$\sigma_1^{X_l} = (\sigma_{c,1}^{X_l} + \sigma_{p,1}^{X_l})/2,$$
$$\sigma_2^{X_l} = (\sigma_{c,2}^{X_l} + \sigma_{p,2}^{X_l})/2,$$
$$\sigma_1^{X_u} = (\sigma_{c,1}^{X_u} + \sigma_{p,1}^{X_u})/2,$$
$$\sigma_2^{X_u} = (\sigma_{c,2}^{X_u} + \sigma_{p,2}^{X_u})/2,$$

which are the average implied volatility associated with the call and put for each exercise price and expiration date respectively. Let

$$\sigma_1 = \sigma_1^{X_l}\left(\frac{X_u - S}{X_u - X_l}\right) + \sigma_1^{X_u}\left(\frac{S - X_l}{X_u - X_l}\right),$$
$$\sigma_2 = \sigma_2^{X_l}\left(\frac{X_u - S}{X_u - X_l}\right) + \sigma_2^{X_u}\left(\frac{S - X_l}{X_u - X_l}\right),$$

be the interpolated implied volatilities for each contract month with weight equal to the ratio of distance between the current stock prices and the chosen near-the-money exercise price respectively.

Finally, the formula for calculating VXO is

$$VXO = \sigma_1\left(\frac{N_{t_2} - 22}{N_{t_2} - N_{t_1}}\right) + \sigma_2\left(\frac{22 - N_{t_1}}{N_{t_2} - N_{t_1}}\right).$$

Here 22 is used rather than 30 in the interpolation since 30 calendar day is equivalent to $30 - 2 * int(30/7) = 22$ trading days. And it is the interpolation of the implied volatility of the nearby and second nearby option series. The weight is the difference of trading days between the trading days left for each contract month and 22 trading days.

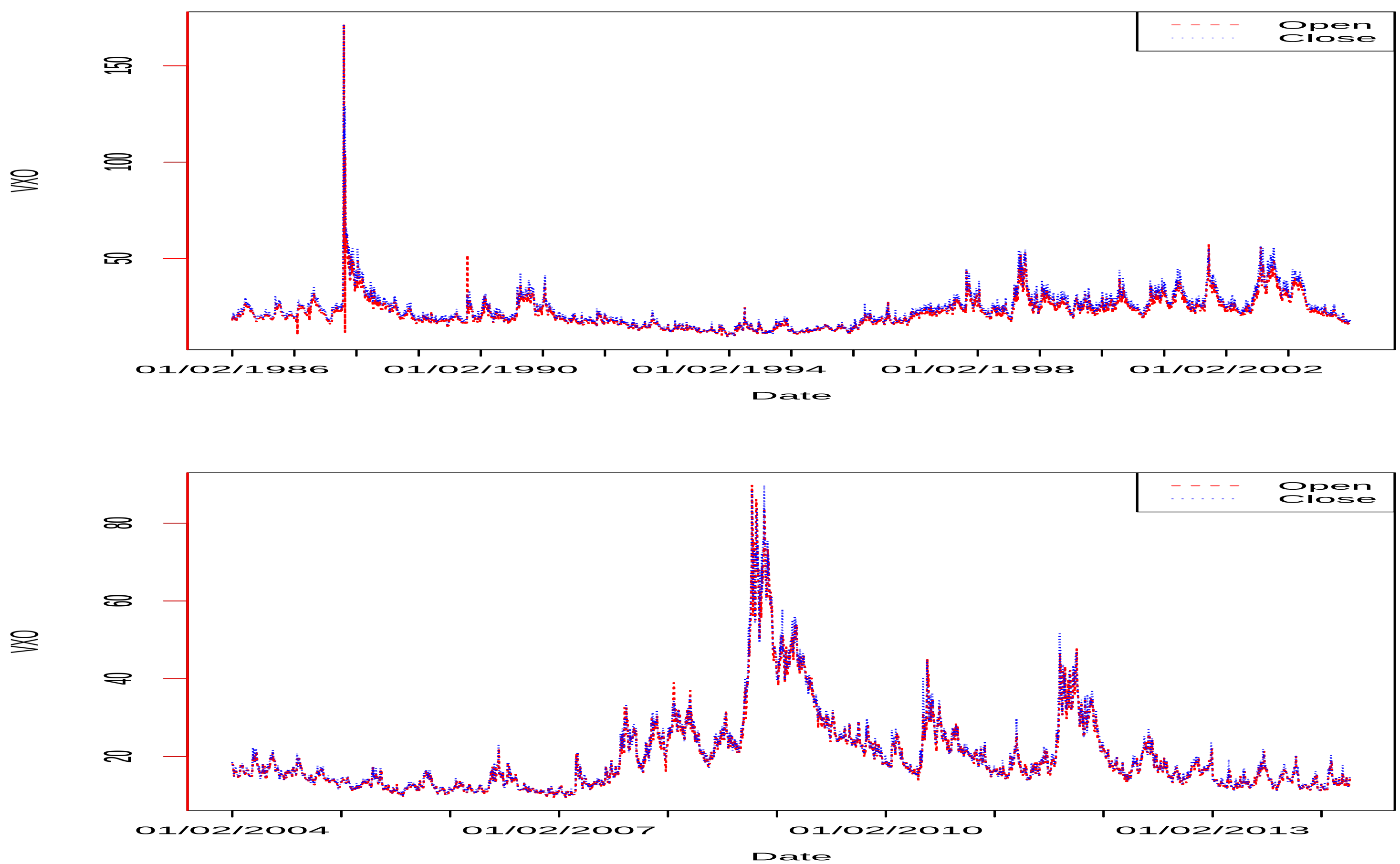
Below are two plots of the historical VXO data[5]. In the plots, the x axis represents the time while the y axis labels the corresponding VXO data. We plot the open volatility with red, close

[4]Here we mean the quoted put/call prices are reported at the same time, hence the calculated volatilities for both put and call are based on the data with same time delay length.

[5]Data available at http://www.cboe.com/micro/vix/historical.aspx.

with light blue, high with green, and low with blue. From the upper plot in Figure 3.1, we can see there is a big spike around 1987. The VXO climted to around 150 when the market crashed on Oct. 1987. After that, the market went back to bull, and the VXO fell to around 20. Then another big spike from the lower plot in Figure 3.1 on Nov. 2008, when global financial crisis exploded due to subprime loans and credit default swaps. We can also see that the intra-day volatility is higher in the opening than in the closing moment.

Figure 3.1 The Trend of the VXO Open and Close Prices for 1986 - Present.



3.1.4 VIX - The Current Volatility Index

The procedures to calculate the VIX is provided in this section. The derivation of the VIX formula will be given in Section 4.2.3.

The trading column of the S & P 500 option market surged to be around thirteen times as large in trading volume as the S & P 100 option market in 2003[6]. Together with the introduction of new estimation method, CBOE had decided to switch from VXO to VIX to measure the market volatility, namely the 30-day volatility of S & P 500 index, on September 22, 2003. The VIX is calculated based on nearby and next nearby European-style option prices based on S & P 500, (ticker symbol SPX). It is required that the nearby option must have at least one week to expire. CBOE provides a short online booklet to show VIX calculation step by step. We introduce the following concepts before presenting the formula. We also provide a very simple example to help grasp the idea. Let time to expiration T be measured according the following rule:

$$M_{current\ \ day} = \text{minutes remaining until midnight of the current day},$$

$$M_{setttlement\ \ day} = \text{minutes from midnight until 8:30a.m. on SPX settlement day},$$

$$M_{Other\ \ days} = \text{total minutes in the days between current day and settlement day},$$

$$N = M_{current\ \ day} + M_{setttlement\ \ day} + M_{Other\ \ days},$$

$$T = N/\text{minutes in 365 days}.$$

Yield of the US T-bill with the closest maturity date for each contract month is adopted as the corresponding risk-free interest rate R. For each contract month, we perform the following procedures. We show it with a data set which is generated by assuming Black-Scholes formula with normal pricing error. Note, in the simulation, expiration is $T = 30/365$ year, diffusion parameter equals 0.2 and the current stock price is 1000. In Table 3.1, we list the stike price (StrikePrice), the bid and ask price for call options ($C - Bid$ and $C - Ask$) and that for the corresponding put options ($P - Bid$ and $P - Ask$).

[6] The daily average trading volume was 145852 contracts in 2003.

Table 3.1 Illustration: The Original Options Prices to Be Selected From.

StrikePrice	C-Bid	C-Ask	P-Bid	P-Ask
825	174.98	175.04	0.00	0.04
850	150.01	150.07	0.01	0.07
875	125.15	125.21	0.15	0.21
900	100.68	100.74	0.00	0.74
925	77.18	77.24	0.00	2.24
950	55.64	55.70	5.64	5.70
960	47.83	47.89	0.00	7.89
975	37.22	37.28	12.22	12.28
1000	22.84	22.90	22.84	22.90
1025	12.74	12.80	37.74	37.80
1050	6.42	6.48	56.42	56.48
1075	0.00	2.97	77.91	77.97
1100	0.00	1.23	101.17	101.23
1125	0.42	0.48	125.42	125.48
1150	0.00	0.18	150.12	150.18
1175	0.02	0.08	175.02	175.08

- ⋆ The average bid/ask price at strike price K_i is quoted as the option prices $Q(K_i)$, which is the mean of the bid and ask prices of each chosen option. For our example, Table 3.2 presents the calculated option prices for both call options and put options.

- ⋆ Find the strike price at which the absolute difference between Call and Put prices are the smallest. Then define the forward index price F according to the put-call parity as

$$F = \text{strick price} + e^{RT} * (\text{call-put}).$$

From Table 3.2, we can see the minimum absolute difference (AbsoluteDifference) for corresponding call prices ($Call$) and put prices (Put) is achieved at strike price 1000. We adopt interest rate $r = 0$. Hence the forward index price is $F = 1000 + e^{0*30/365} * (22.87 - 22.87) = 1000$.

- ⋆ Choose K_0 to be the strike price immediately below or equal to the forward index price F, and treat it as the center. In our example, the forward price is equal to the strike price, which is not usually in real world, because we used simulated exact data. Hence, K_0 is chosen to 1000.

Table 3.2 Illustration: Seeking the Smallest Absolute Difference of the Call and the Put Prices.

StrikePrice	Call	Put	AbsoluteDifference
825	175.01	0.02	174.99
850	150.04	0.04	150.00
875	125.18	0.18	125.00
900	100.71	0.37	100.34
925	77.21	1.12	76.09
950	55.67	5.67	50.00
960	47.86	3.95	43.92
975	37.25	12.25	25.00
1000	22.87	22.87	0.00
1025	12.77	37.77	25.00
1050	6.45	56.45	50.00
1075	1.48	77.94	76.45
1100	0.62	101.20	100.59
1125	0.45	125.45	125.00
1150	0.09	150.15	150.06
1175	0.05	175.05	175.00

⋆ Start from the center K_0, move successively to the two ends of more extreme strike prices respectively, choose out-of-the-money put options with strike price $K_j \leq K_0$ and out-of-the-money call options with strike price $K_j \geq K_0$. Here, an out-of-the-money call option means the strike price is no smaller than the current stock price, and an out-of-the-money put option is with strike price no greater than the current stock price. Exclude any option with a bid price 0 and any option further away from the center with two consecutive zero bid nearer than it to the center. In our example, as illustrated in Table 3.3, for put options, we start from put option with strike price 1000, then move to those with lower option prices. We can see when strike prices are $960, 925, 900, 825$, the corresponding bid price is 0. So those options will not be included. Because the bid price for the strike price 925 and 900 are consecutively 0, any option with strike price lower than 900 will be excluded too. Only put options with strike price $1000, 975, 950$ will be included for further calculation. Similarly for call options, we start from call option with strike price 1000, then move to those with strike prices larger than 1000. Delete options with strike price $1175, 1100, 1150$ and those with strike price larger than 1100, because two consecutive 0 bids are observed at strike price $1075, 1100$. The call options to be further considered are with strike prices $1000, 1025$ and 1050.

Table 3.3 Illustration: Exclude Invalid Option Prices.

StrikePrice	C-Bid	C-Ask	P-Bid	P-Ask
825	174.98	175.04	0.00	0.04
850	150.01	150.07	0.01	0.07
875	125.15	125.21	0.15	0.21
900	100.68	100.74	0.00	0.74
925	77.18	77.24	0.00	2.24
950	55.64	55.70	5.64	5.70
960	47.83	47.89	0.00	7.89
975	37.22	37.28	12.22	12.28
1000	22.84	22.90	22.84	22.90
1025	12.74	12.80	37.74	37.80
1050	6.42	6.48	56.42	56.48
1075	0.00	2.97	77.91	77.97
1100	0.00	1.23	101.17	101.23
1125	0.42	0.48	125.42	125.48
1150	0.00	0.18	150.12	150.18
1175	0.02	0.08	175.02	175.08

The selected call and put options are summarized in Table 3.4.

Table 3.4 Illustration: The Selected Put and Call Option Prices.

StrikePrice	Put
950	5.67
975	12.25
1000	22.87

StrikePrice	Call
1000	22.87
1025	12.77
1050	6.45

⋆ The K_0 put and call prices are averaged to produce a single value. In our example, see Table 3.4, we can find the put option price and call option price with strike price 1000 are both 22.87. Hence, further average them we get a single value to be used as option price with strike price 1000, ignoring the option type:

$$(22.87+22.87)/2=22.87.$$

⋆ Define

$$\triangle K_j = |K_{j+1}-K_{j-1}|/2,$$

if $j-th$ strike price is one of the chosen strike prices other than the maximal and minimum one. If $m-th$ strike price is the chosen strike price with the minimum value and the maximal chosen strike price is the $n-th$, then define

$$\triangle K_m = (K_{m-1}-K_m), \text{ and } \triangle K_n = (K_n - K_{n-1}).$$

The values of $\triangle K$ for our case is listed in Table 3.5. For the minimal strike price 950, the corresponding $\triangle K$ is $975-950=25$. The $\triangle K$ for the maximal strike price is $1050-1025=25$. For strike price 975, it is neither the minimal nor the maximal strike price, so the $\triangle K$ equals $|1000-950|/2=25$. Similarly, the value of $\triangle K$ corresponds to strike prices 1000 and 1025 are both 25.

⋆ The expected volatility of the underlying asset for the future T year, denoted as σ, is calculated as

$$\hat{\sigma}^2 = \frac{2}{T}\sum_{j\in J}\frac{\triangle K_j}{K_j^2}e^{RT}Q(K_j) - \frac{1}{T}[\frac{F}{K_0}-1]^2. \tag{1.18}$$

Here $J=\{j : K_j \text{ is the j-th chosen strike price}\}$. Derivation of this model free estimator will be shown in Section 3.2.3. Table 3.5 consists the merged chosen options with corresponding strike price, option price, $\triangle K$, and contribution of each individual option ($Contribution$) to the $\hat{\sigma}^2$, which equals $\frac{2}{T}\frac{\triangle K_j}{K_j^2}e^{RT}Q(K_j)$. When we sum up the contribution of each option to the VIX, we get 0.03652708. The adjustment to approximation error part for our idealized data turns out to be $\frac{1}{T}[\frac{F}{K_0}-1]^2 = \frac{1}{30/365}[\frac{1000}{1000}-1]=0$. Hence,

$$\hat{\sigma}^2 = 0.03652708-0=0.03652708,$$

which represents the expected future $30/365-$year volatility of the underlying stock.

Table 3.5 Illustration: The Individual Contribution.

StrikePrice	OptionPrice	DeltaK	Contribution
950	5.67	25	0.003820090
975	12.25	25	0.007840117
1000	22.87	25	0.013913500
1025	12.77	25	0.007394941
1050	6.45	25	0.003558428

Now, we proceed to define VIX as the interpolation of the volatility σ calculated by the above method for each contract month. Let $\hat{\sigma}_1$ and $\hat{\sigma}_2$ be the corresponding volatility for the nearby and next nearby contract month calculated by (1.18). Denote the time to the nearby option by N_{T_1}, and the time to expire for the next nearby option to be N_{T_2}[7]. Write $T_1 = N_{T_1}/(60*24*365)$ and $T_2 = N_{T_2}/(60*24*365)$. N_{30} and N_{365} are the total calendar minutes in 30 days and 360 days respectively. Then the 30-day volatility index, VIX, is given below through interpolation:

$$VIX = 100\sqrt{\left[T_1\hat{\sigma}_1^2\left(\frac{N_{T_2}-N_{30}}{N_{T_2}-N_{T_1}}\right)+T_2\hat{\sigma}_2^2\left(\frac{N_{30}-N_{T_1}}{N_{T_2}-N_{T_1}}\right)\right]*\frac{N_{365}}{N_{30}}}, \tag{1.19}$$

In our example, we only have one contract month which happen to equal 30 days. Hence, no further weighting using (1.19) is needed. The estimated 30-day VIX be $\sqrt{0.03652708-0}*100 = 19.11206$, compared with our true $VIX = 100*\sigma = 100*0.2 = 20$. However, if we have two different contract months, say in 15 and 45 days, then $N_{T_1} = 60*24*15$, $N_{T_2} = 60*24*45$, $T_1 = \frac{60*24*15}{60*24*365} = \frac{15}{365}$, $T_2 = \frac{45}{365}$. $\hat{\sigma}_1^2 = 0.042$ and $\hat{\sigma}_2^2 = 0.030$. Plug these value into (1.19), estimated VIX value is

$$100*\sqrt{\left\{\frac{15}{365}*0.042*\frac{64800-43200}{64800-21600}+\frac{45}{365}*0.030*\frac{43200-21600}{64800-21600}\right\}*\frac{525600}{43200}} = 18.1659.$$

Two plots for the historical VIX data[8] are provided in Figure 3.2 and 3.3. In the plots, the x axis represents the time, the left side y axis labels the value of the corresponding VIX data, and the right side y axis stands for the corresponding same day value of S & P 500 index. We

[7]Their definitions are provided at the beginning of this subsection.

[8]Historical VIX data can be downloaded from http://www.cboe.com/micro/vix/historical.aspx. SPX data is available from http://finance.yahoo.com/q/hp?s=^GSPC+Historical+Prices.

plot the open VIX in red dashed line with round points, and the SPX in blue dotted line with blue stars.

Figure 3.2 shows the historical trend VIX and SPX during 01/02/1990 and 04/08/2014. The VIX curve tend to go up while the SPX curve go down, and vice verse. And we see the VIX spikes with much higher rate when SPX falls than when it rises. The later intuitive conclusion will be justified by a regression analysis in Section 4.1.6.

In Figure 3.3, we zoom in to see more details. The upper left plot shows the trend of the VIX and the SPX during a relatively peaceful market period between 12/01/1992 and 06/06/1993. Both the VIX and the SPX curve goes up and down in relatively small ranges, say 11 to 16 for VIX and 420 to 460 for SPX. The stock market went through a very steady growth period since 1992 till Sep. 2000. Investors are very confident in that bullish market and SPX climbed from around 500 to over 1500. In the upper right plot, during 07/02/2000 and 10/12/2000, we can see an obvious convex curve for SPX and a concave one for VIX just around the period of the big crash of the stock market on September 2000. Generally speaking, the VIX were going down to as low as 16 and the SPX went up sharply to reach 1520.77 on 09/06/2009. In the lower left plot, we see an opposite trend during 08/02/2008 and 06/03/2009. In 2008, the stock market was hit hard by the subprime mortgage crisis, which was caused by the final exposure of a set of financial problems and lead to worldwide manifest recession. During the last quarter of 2008, the USA Federal Reserve, the European Central Bank, and other central banks made the largest monetary injection into the market in world history to save the market. The market started to pick up confidence in the spring of 2009. Hence, we see in the plot a big concave curve for the SPX, while the VIX went through a shocking increase of 20 to 80 in four months and come down to 20 slowly not until the end of 2009. The lower right plot illustrate the dynamic of SPX and VIX during 10/02/2013 and 04/08/2014. For the last half year, we are experiencing a relatively peacefully financial market with no major crisis, except for the short sells off caused by Ukraine-Russia tension around March 2014. Correspondingly, we see a big spike of VIX around that time period.

Figure 3.2 The Trend of the VIX Data versus the S & P 500 Data for 1990 - Present.

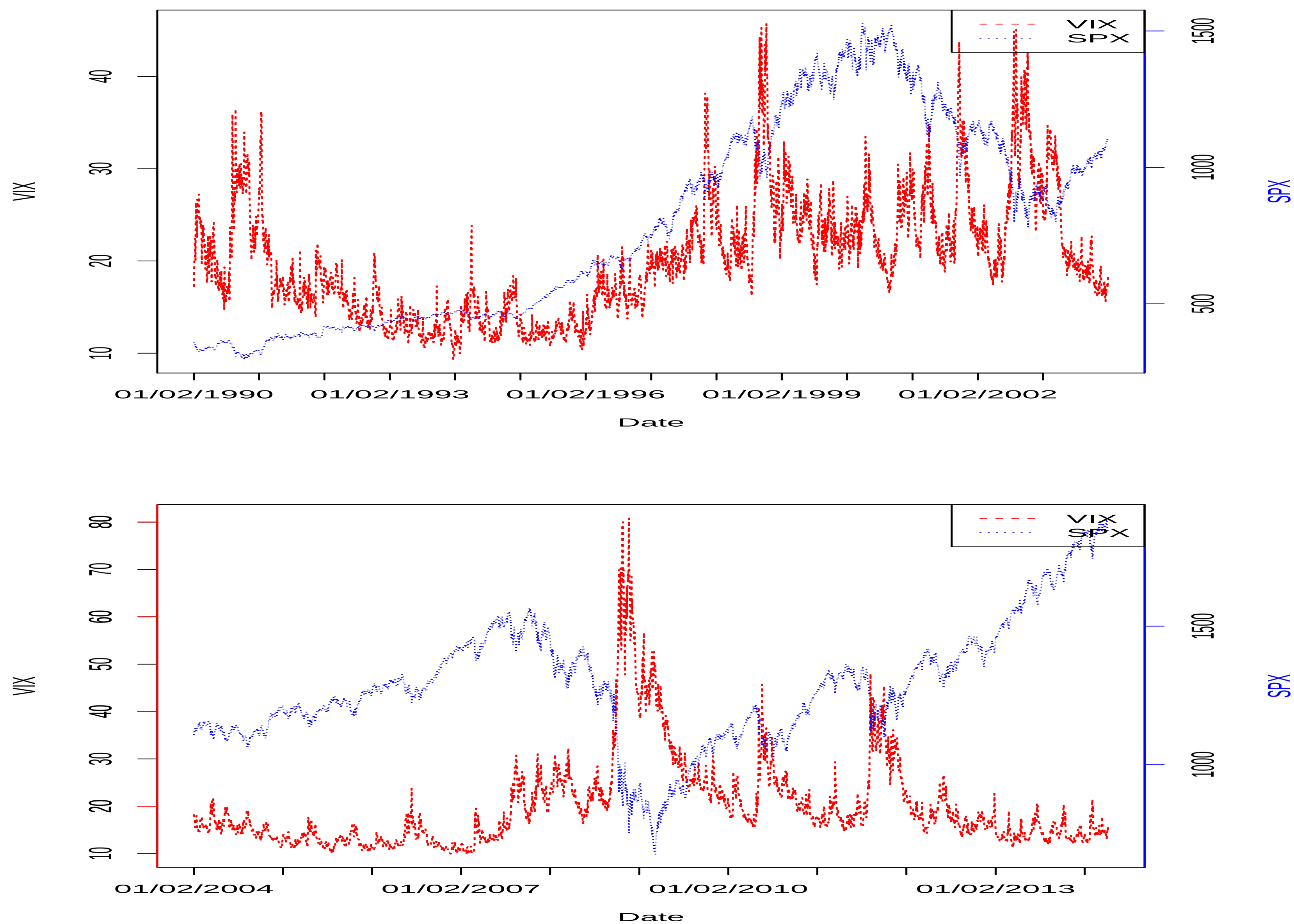
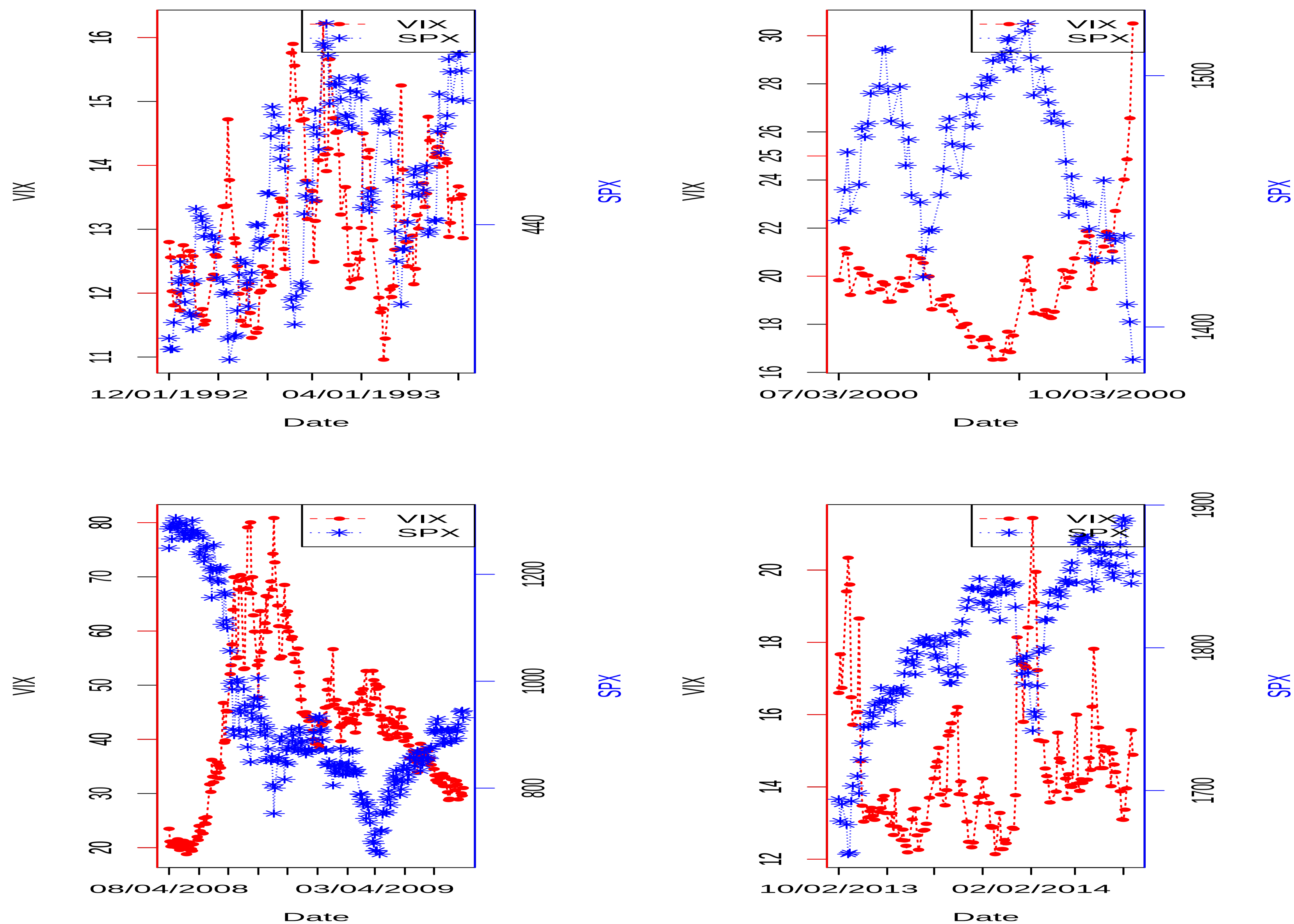


Figure 3.3 The Trend of the VIX Data versus the S & P 500 Data for Selected Months.



3.1.5 Discussion on VXO and VIX

We briefly discuss the relation between VXO and VIX below. We also note some properties of VIX.

VXO uses eight near-the-money options, four of each coming from the nearby or the second nearby OEX options series. VIX includes some selected out-of-the-money options from the near-term and next-term SPX options. VXO is defined as weighted sum of implied volatility, which can be calculated by the Cox-Ross-Rubinstein method (1979), i.e. binary tree method. VIX is defined to be the weighted sum of integrated volatility, which can be estimated by the method provided in CBOE white paper (2009). The estimated integrated volatility is essentially weighted sum of option prices with a range of strike prices.

To see the general range and trend of VXO and VIX, we provide the five number summary for VXO dada (1986-2003) and VIX data (2004-2014) in Table 3.6. We also plot these yearly medians, minimal, and maximal value of VXO/VIX in the upper plot in Figure 3.4. From the plot, we can easily see the normal range of VXO and VIX is around 20, between 0 and 40. The two largest value observed are 150.20 in 1987 and 80.86 in 2008.

Intuitively, we see the value of VXO and VIX are very similar from the lower plot in Figure 3.4. Similarly as in Whaley (2009), we explore how to approximately estimate the value of VIX when only VXO is available. The daily standard deviation of the S & P 100 was 0.011822. The S & P 500 volatility is only 96.24 of the S and P 100 volatility. The simple linear regression without intercept of the VIX based on SPX w.r.t. VIX based on OEX range from 1/2/2004 to 9/24/2012 is with slope 0.987697. The standard deviation for the estimation is 0.001337, with $R^2 = 0.996$. Hence a simple estimation of VIX based on SPX prior to the new method can be

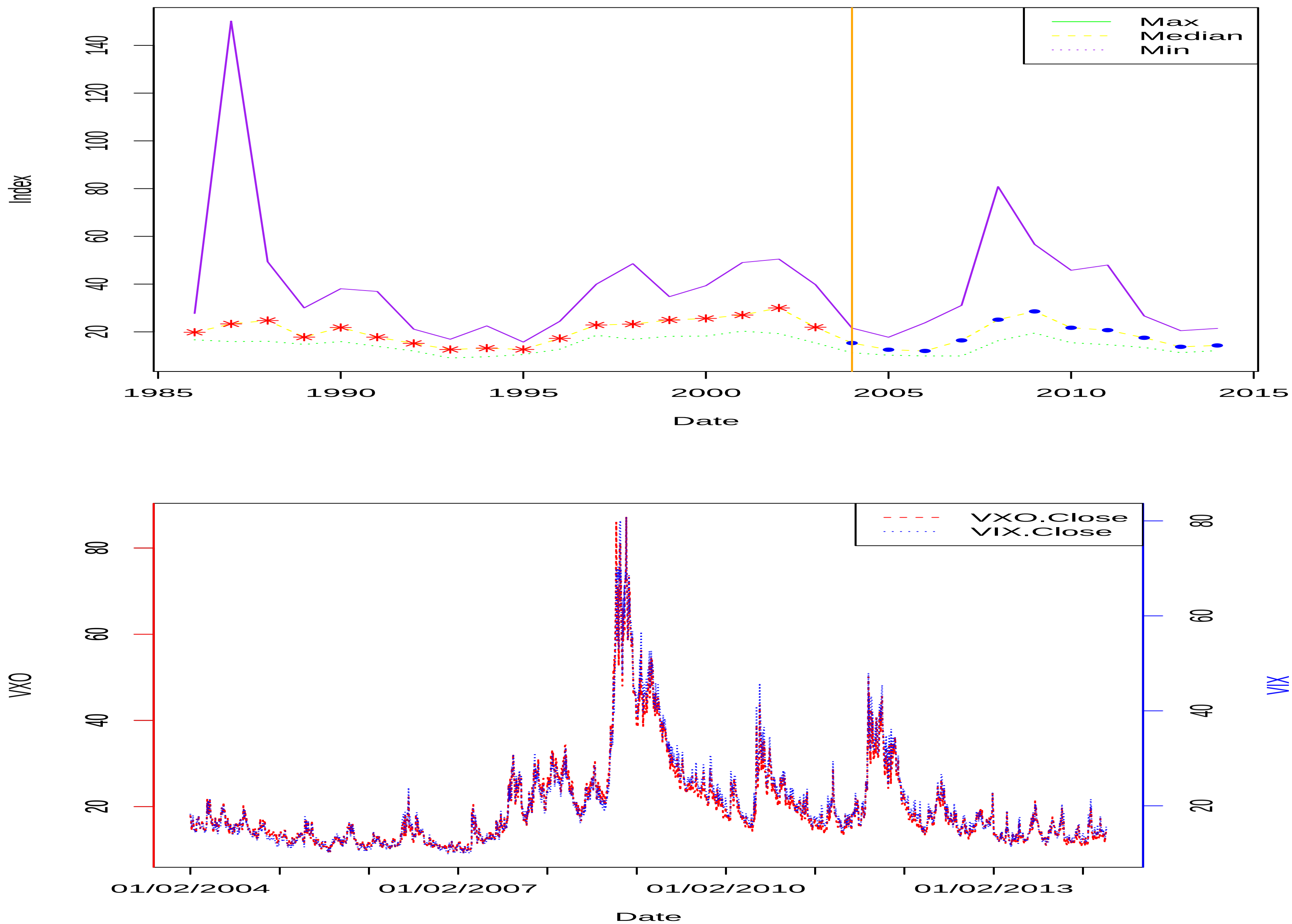
$$V\hat{I}X_{SPX} = 0.987697 * VIX_{OEX}.$$

We observe that the VIX spikes when the market collapses. This is why VIX is called fear index. And we expect the change in VIX rises at a higher absolute rate when the stock market

Table 3.6 Summary Statistics for the VXO Data (1986-2003) and the VIX Data (2004-2014) by Year.

Year	Num.Days	Min.	1stQu.	Median	Mean	3rdQu.	Max.
1986	252	16.66	18.58	19.82	20.41	21.66	27.69
1987	253	15.91	21.46	23.33	29.28	27.59	150.20
1988	253	16.06	20.97	24.75	25.48	28.01	49.36
1989	252	14.76	16.94	17.79	18.34	18.73	30.03
1990	253	15.92	18.84	21.82	22.96	26.87	38.07
1991	251	13.93	16.48	17.77	18.74	19.68	36.93
1992	254	11.98	13.73	15.17	15.27	16.43	21.12
1993	251	9.04	11.72	12.61	12.65	13.40	16.90
1994	252	9.59	11.62	13.18	13.37	14.89	22.50
1995	252	10.49	11.84	12.63	12.70	13.55	15.72
1996	253	12.74	16.16	17.25	17.43	18.66	24.43
1997	253	18.55	21.72	22.87	24.03	25.34	39.96
1998	251	16.88	21.03	23.24	25.93	28.46	48.56
1999	252	18.13	23.06	24.96	25.26	27.34	34.74
2000	252	18.23	23.13	25.62	25.93	28.40	39.33
2001	248	20.29	24.52	27.04	28.36	31.61	49.04
2002	251	19.25	23.10	30.01	30.58	36.30	50.48
2003	252	15.35	19.57	21.94	24.08	27.72	39.77
2004	252	11.23	14.30	15.32	15.48	16.55	21.58
2005	252	10.23	11.68	12.52	12.81	13.64	17.74
2006	251	9.90	11.36	12.00	12.81	13.62	23.81
2007	251	9.89	13.13	16.43	17.54	21.66	31.09
2008	253	16.30	21.58	25.10	32.69	40.00	80.86
2009	252	19.47	24.28	28.57	31.48	39.31	56.65
2010	252	15.45	18.34	21.72	22.55	25.20	45.79
2011	252	14.62	17.40	20.72	24.20	31.56	48.00
2012	250	13.45	15.75	17.52	17.80	19.05	26.66
2013	252	11.30	12.98	13.74	14.23	14.98	20.49
2014	67	12.14	13.72	14.30	14.75	15.27	21.44

Figure 3.4 Plot the Yealy Summary Statistics of the VXO Data (1986-2003) and the VIX Data (2004-2014) and Contrast Plot for the Trend of the VXO and the VIX since 2004.



falls than when it rises. Whaley (2009) verified this by the following regression. If we regress the daily percentage rate of change of VIX, denoted as $RVIX_t$, on the percentage rate of change of the S & P 500 portfolio, write as $RSPX_t$, and the rate of change of the S & P 500 portfolio conditional on the market going down and 0 otherwise, defined as $RSPX_t^-$, we can see:

$$RVIX_t = -0.004 - 2.990 RSPX_t - 1.503 RSPX_t^-,$$

which is consistent with the expectation that the increased demand to buy index puts would result in a higher absolute rate rise for VIX when the stock market falls than when it rises.

The VIX calculation method provide an easy practical way for trading volatility. CBOE introduced VIX future in 2004 and VIX option in 2006. Less than five years later, the average daily transaction for both the VIX options and futures has reached 0.1 million contracts.

3.2 Theory

3.2.1 Existing Works

The original VIX, or later called VXO, was introduced by Waley in 1993. Later, based on the evolution of the market and the work by Demeterfi, Derman, Kamal and Zou (1999), CBOE switched from VXO to the VIX to include out-of-the-money options and change to more popular S & P 500 option market. CBOE white paper (2009) provides detailed procedures to perform the VIX calculation when finite number of options with a range of strike prices are available. Jiang and Tian (2007) studied the approximation error introduced by the method in CBOE white paper (2009). They demonstrated how large these errors can be by simulated data under Black-Scholes model. They suggested to decrease the errors by using cubic spline method. They provided simulation results under the stochastic volatility with jump model.

Britten-Johns and Neuberger (2000) proposed a model-free method under diffusion assumption to forecast realized volatility with only call options:

$$E_0[\int_0^{t_2}(\frac{dS_t}{S_t})^2] = 2\int_0^{\infty}\frac{C(t_2,K)-max(S_0-K,0)}{K^2}dK.$$

Jiang and Tian (2005) extended their results to model-free implied volatility with jumps for asset price. Song and Xiu (2012) considered the non-parametric estimation of state-price densities and conclude that the state-of-the-art stochastic volatility models in the literature cannot capture the S & P 500 and VIX option prices simultaneously.

In this section, we first introduce the kernel regression method and the derivation of the VIX formula. To reduce estimation error suggested by CBOE method when option pricing error presents, a method that combines kernel smooth into the CBOE procedure is suggested. Simulation under Black-Schole model and Merton model with jump supports our error deduction approach.

3.2.2 N-W Estimator

Giving a data set, we may want to learn the logic behind the values. When no underlying model is clearly known to represent the relation, we may try to let the data to speak for themselves. This is what some non-parametric analysis and references can do. In this subsection, we briefly introduce one kind of non-parametric regression method, N-W estimator. We define the concept of kernel, specify the N-W estimator, and describe three different bandwidth selection method.

Let K be a function such that

$$\int_{-\infty}^{\infty}K(x)dx = 1,$$
$$\int_{-\infty}^{\infty}xK(x)dx = 0,$$
$$\int_{-\infty}^{\infty}x^2K(x)dx = \sigma_K^2 > 0.$$

K is called a kernel function, which is usually a symmetric probability distribution function itself. Some commonly used kernel functions are

Gaussian kernel:	$K(x) = \frac{1}{2\pi}e^{-x^2/2}$,
Epanechnikov kernel:	$K(x)\frac{3}{4}(1-x^2)\infty_{\lvert x\rvert<1}$,
Biweight kernel:	$K(x) = \frac{15}{16}(1-x^2)^2\infty_{\lvert x\rvert<1}$.

The Gaussian kernel is well adapted because it is with many well known good properties that come in simple forms. Both the Epanechnikov kernel and the Biweight kernel are with compact support, and the later is smoother than the former.

Suppose we have a random sample $\{X_i, Y_i\}_{1\le i\le n}$. The model under consideration is

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i,$$

where $E(\varepsilon|X = x) = 0$, $Var(\varepsilon|X = x) = 1$, $m(x) = E(Y|X = x)$ is the unknown conditional mean function, $\sigma^2(x) = Var(Y|X = x)$ is the unknown conditional variance function. Motivated by the fact that

$$\begin{aligned} m(x) &= E(Y|X = x) \\ &= \int yf(y|x)dy \\ &= \int y\frac{f(x,y)}{f(x)}dy, \end{aligned}$$

where $f(x)$, $f(x,y)$ and $f(y|x)$ are the true probability distribution function for X, (X,Y) and $Y|X$, Nadaraya (1963) and Watson (1964) proposed the N-W estimator for the regression:

$$\begin{aligned} \hat{m}(x) &= \int y\hat{f}(y|x)dy \\ &= \frac{\sum_{i=1}^n K(\frac{x-X_i}{h})Y_i}{\sum_{i=1}^n K(\frac{x-X_i}{h})}. \end{aligned}$$

where h is the bandwidth and the K is a kernel function. The bias and variance for the estimator are

$$\begin{aligned} Bias(\hat{m}(x)) &= \frac{1}{2}h^2\frac{m^{"}(x)f(x) + 2m^{'}(x)f^{'}(x)}{f(x)}\sigma_K^2 + o(h^2) + O(\frac{1}{nh}), \\ Var(\hat{m}(x)) &= \frac{\sigma^2(x)R(K)}{nhf(x)} + o(\frac{1}{nh}), \end{aligned}$$

where $R(K) = \int K^2(t)dt$.

Various values can be adapted for the bandwidth, h, depending on the sample values. We introduce three types of bandwidth selection method below.

The optimal global bandwidth minimizes the mean square error of the N-W estimator. The mean square error of the N-W estimator is

$$MSE\{\hat{m}(x)\} = \frac{1}{4}h^4 b^2(x)\sigma_K^4 + \frac{\sigma^2(x)R(K)}{nhf(x)} + o\{(nh)^{-1} + h^4\},$$

where $b(x) = \{m^{''}(x) + 2m^{'}(x)f^{'}(x)\}/f(x)$. The corresponding optimal bandwidth which minimize it is

$$h_{opt} = (\frac{\sigma^2(x)}{b^2(x)f(x)})^{1/5}(\frac{R(K)}{\sigma_K^4})^{1/5}n^{-1/5},$$

and the minimal MSE achieved under it is

$$MSE_{opt}(\hat{m}(x)) = \frac{5}{4}\sigma^{8/5}(x)b^{2/5}(x)f^{-4/5}(x)(R(K)\sigma_K^{-1})^{4/5}n^{-4/5}.$$

Note that as h_{opt} depends on $m^{'}(x)$, $m^{''}(x)$, $f^{'}(x)$, $\sigma^2(x)$ and $f(x)$, it is hard to use the plug-in method to get estimation. And hence some other method like the following two might be better choices for practical use.

Both the cross-validation method and the penalizing function method seek to minimize the average square error. In the cross-validation method, the target is to find a h that minimizes

$$CV(h) = \frac{1}{n}\sum_{j=1}^{n}\{Y_j - \hat{m}_{h,j}(X_j)\}^2,$$

where $\hat{m}_{h,j}(x) = \{\sum_{i \neq j} K(\frac{x-X_i}{h})Y_i\}/\{\sum_{i \neq j} K(\frac{x-X_i}{h})\}$. And the penalizing function method aims to find a h to minimize

$$G(h) = n^{-1}\sum_{j=1}^{n}(Y_j - \hat{m}_h(X_j))^2\beta(W_j(X_j)),$$

where $\beta(u) = 1 + 2u$ is a penalty function proposed by Shibata (1981).

Please refer to Härdle (1989) for more detailed introduction about kernel regression.

3.2.3 Derivation of the VIX formula

The VIX formula given in Section 4.1.5 is derived based on Demeterfi, Derman, Kamal and Zou (1999). It does not come as naturally as the VXO. Hence we provide a derivation of VIX Formula 1.19 to help understand the internal logic. Most related theorems are recited in Section 4.1.2. For more details, please refer to Øksendal (1998), Karatzas and Shreve (1988), and Lamberton, D. and Lapeyre, B. (1996).

Suppose Stock price $\{S_t\}$ follows a general diffusion process

$$\frac{dS_t}{S_t} = \mu(t;S_t,\theta)dt + \sigma(t;S_t,\theta)dB_t, \tag{2.20}$$

which further, by Itô formula, indicates,

$$\begin{aligned} d(ln(S_t)) &= \frac{dS_t}{S_t} + \frac{1}{2}(-\frac{1}{S_t^2})(dS_t)^2 \\ &= \frac{dS_t}{S_t} - \frac{1}{2}\frac{1}{S_t^2}(\sigma(t;S_t,\theta)S_t)^2 tdt \\ &= \frac{dS_t}{S_t} - \frac{1}{2}\sigma^2(t;S_t,\theta)dt. \end{aligned}$$

The realized variance is

$$V = \frac{1}{T}\int_0^T \sigma^2(t;S_t,\theta)dt,$$

the fair delivery price of future realized variance is

$$\begin{aligned} K_{var} &= E[V] = E[\frac{1}{T}\int_0^T \sigma^2(t;S_t,\theta)dt] \\ &= E[\frac{1}{T}\int_0^T 2\Big(\frac{dS_t}{S_t} - d(ln(S_t))\Big)] \\ &= \frac{2}{T}E[\int_0^T \frac{dS_t}{S_t} - ln(\frac{S_T}{S_0})]. \end{aligned} \tag{2.21}$$

The ultimate purpose is to convert the expression of equation (2.21) to a form that involves only simple observable asset prices. In order to realize it, we now focus on the calculation of the above expectation. Suppose the interest rate is constant, write as $R > 0$. In a fair market, the discounted stock price

$$\tilde{S}_t = e^{-Rt}S_t,$$

which represent the current value of that stock, should be a martingale. Note

$$\begin{aligned} d\tilde{S}_t &= d(e^{-Rt}S_t) = -re^{-Rt}S_t dt + e^{-Rt}dS_t \\ &= -re^{-Rt}S_t dt + e^{-Rt}S_t\mu(t;S_t,\theta)dt + e^{-rt}S_t\sigma(t;S_t,\theta)dB_t \\ &= -r\tilde{S}_t dt + \tilde{S}_t\mu(t;S_t,\theta)dt + \tilde{S}_t\sigma(t;S_t,\theta)dB_t \\ &= \tilde{S}_t\sigma(t;S_t,\theta)\Big(dB_t + \frac{\mu(t;S_t,\theta)-R}{\sigma(t;S_t,\theta)}dt\Big). \end{aligned}$$

Suppose $\sigma(t;S_t,\theta) \neq 0$ almost surely. Define $\theta_t = B_t + \frac{\mu(t;S_t,\theta)-R}{\sigma(t;S_t,\theta)}$, and

$$L_t = \exp\{-\int_0^t \theta_s dB_s - \frac{1}{2}\int_0^t \theta_s^2 ds\}.$$

Assume

- ⋆ $\{\theta_t\}_{0\le t\le T}$ is an adapted process satisfying $\int_0^t \theta_s^2 ds < \infty$,
- ⋆ $E[\exp(\frac{1}{2}\int_0^T \theta_t^2 dt)] < \infty$,

then $\{L_t\}_{0\le t\le T}$ is a martingale. Then under probability P^L with density L_T relative to P, the process $\{B_t^*\}_{0\le t\le T}$ defined by

$$B_t^* = B_t + \int_0^t \theta_s ds$$

is a standard Brownian motion.

Therefore, process in (2.20) can be rewritten as

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu(t;S_t,\theta)dt + \sigma(t;S_t,\theta)(dB*_t - \theta_t dt) \\ &= \sigma(t;S_t,\theta)dB*_t + \Big(\mu(t;S_t,\theta) - \sigma(t;S_t,\theta)\frac{\mu(t;S_t,\theta)-R}{\sigma(t;S_t,\theta)}\Big)dt \\ &= \sigma(t;S_t,\theta)dB*_t + Rdt. \end{aligned}$$

Denote the expectation under probability measure P^L as E^*. Suppose $\{\sigma(t;S_t,\theta)\}_{0\le t\le T}$ is measurable and adapted in the new equivalent probability space, and $E^*[\int_0^T \sigma^2(t;S_t,\theta)dt] < \infty$. Then, inder the risk neutral probability P^L,

$$E^*[\int_0^T \frac{dS_t}{S_t}] = RT. \tag{2.22}$$

Now we work on the expectation of the second term for any positive number K_0,

$$ln(\frac{S_T}{S_0}) = ln(\frac{S_T}{K_0}) + ln(\frac{K_0}{S_0}), \tag{2.23}$$

by Taylor expansion,

$$-ln(\frac{S_T}{K_0}) = -\frac{S_T - K_0}{K_0} + \int_0^{K_0} \frac{1}{K^2} max(K - S_T, 0) dK + \int_{K_0}^{\infty} \frac{1}{K^2} max(S_T - K, 0) dK. \tag{2.24}$$

Therefor, combine results from (2.21), (2.22), (2.23), and (2.24), we get

$$\begin{aligned} \frac{T}{2} K_{var} &= RT - ln(\frac{K_0}{S_0}) + E^*\Big[-\frac{S_T - K_0}{K_0} && (2.25) \\ &+ \int_0^{K_0} \frac{1}{K^2} max(K - S_T, 0) dK + \int_{K_0}^{\infty} \frac{1}{K^2} max(S_T - K, 0) dK\Big]. && (2.26) \end{aligned}$$

In the risk-neutral market the forward price $F = e^{RT} S_0$. Due to the martingale property of the discounted price, $E^*[S_T] = e^{RT} S_0$. Recall the definition of the call and put option prices at expiration, we have

$$\begin{aligned} \frac{T}{2} K_{var} &= ln(\frac{F}{S_0}) - ln(\frac{K_0}{S_0}) - [\frac{e^{RT} S_0}{K_0} - 1] \qquad (2.27) \\ &+ \int_0^{K_0} \frac{1}{K^2} e^{RT} P(K,T) dK + \int_{K_0}^{\infty} \frac{1}{K^2} e^{RT} C(K,T) dK \\ &= ln(\frac{F}{K_0}) - [\frac{F}{K_0} - 1] \\ &+ \int_0^{K_0} \frac{1}{K^2} e^{RT} P(K,T) dK + \int_{K_0}^{\infty} \frac{1}{K^2} e^{RT} C(K,T) dK. \end{aligned}$$

Applying Taylor expansion to the second term in the last step of (2.27), truncate the integration to a finite interval $[K_L, K_U]$, and adopt the discrete approximation of integrals by their Riemann Sums, we arrive at

$$\frac{T}{2} K_{var} \approx -\frac{1}{2}[\frac{F}{K_0} - 1]^2 + \sum_{j \in J} \frac{\triangle K_j}{K_j^2} e^{RT} Q(K_j). \tag{2.28}$$

Here $J = \{j : K_j \text{ is the j-th chosen strike price}\}$. This implies the validity of (1.18), which is used to calculate the VIX. The term $-\frac{1}{2}[\frac{F}{K_0} - 1]^2$ is an adjustment compensation for the fact that the option series is not centered around a strike exactly at-the-money. All the terms in the final form are easy to be harvested from the market. The second term is a weighted sum of the near-the-money option prices for both call and put options.

3.2.4 VIX Estimation with Kernel Smooth

The final VIX formula (1.18) adopted by CBOE is subject to expansion, truncation, and discretization errors. The expansion error is caused by the following replacing, which was justified by Taylor expansion:

$$ln(\frac{F}{K_0}) \approx [\frac{F}{K_0} - 1] - \frac{1}{2}[\frac{F}{K_0} - 1]^2. \tag{2.29}$$

The truncation error is introduced when we cut the infinite range of strike prices with a finite selected range $[K_L, K_U]$:

$$\begin{aligned} & \int_0^{K_0} \frac{1}{K^2} e^{RT} P(K,T) dK + \int_{K_0}^{\infty} \frac{1}{K^2} e^{RT} C(K,T) dK \\ \approx \;& \int_{K_L}^{K_0} \frac{1}{K^2} e^{RT} P(K,T) dK + \int_{K_0}^{K_U} \frac{1}{K^2} e^{RT} C(K,T) dK \end{aligned} \tag{2.30}$$

The discretization error is produced when we do the further approximate

$$\begin{aligned} & \int_{K_L}^{K_0} \frac{1}{K^2} e^{RT} P(K,T) dK + \int_{K_0}^{K_U} \frac{1}{K^2} e^{RT} C(K,T) dK \\ \approx \;& \sum_{j \in J} \frac{\triangle K_j}{K_j^2} e^{RT} Q(K_j) \end{aligned} \tag{2.31}$$

We concentrate on reducing the discretization errors, as well as reducing the bias introduced by the option pricing error, by using the kernel smoothing method.

Suppose the observed call option prices $\{\tilde{C}_i\}$ and put option prices $\{\tilde{P}_j\}$, which are written as $\tilde{Q}$ below, are generated by the following model:

$$\tilde{Q}_i = Q(S, K_i, T, \Sigma, R) + \sigma(Q_i)\varepsilon_i,$$

where $E(\varepsilon|S, K_i, T, \sigma, R) = 0$, $Var(\varepsilon|S, K_i, T, \sigma, R) = 1$, $Q_i = Q(S, K_i, T, \sigma, R) = E[\tilde{Q}|S, K_i, T, \sigma, R]$ is the unknown true theoretical option price, and $\sigma^2 = Var(\tilde{Q}|S, K_i, T, \sigma, R)$ is the unknown conditional variance function.

First, we follow the CBOE method to select those option prices $\{\tilde{Q}_i\}$ that will be used in the final calculation. Then we perform a kernel smooth to those option prices. The estimated option prices $\{\hat{Q}_i\}$, or denoted as $\hat{Q}$ below, are with the form

$$\hat{Q}(x) \quad = \quad \frac{\sum_m K(\frac{x-K_m}{h})\tilde{Q}_m}{\sum_m K(\frac{x-K_m}{h})}.$$

Then we plug in our estimated option prices $\{\hat{Q}_i\}$ at more strike prices along the regression line, into the CBOE VIX pricing formula:

$$\hat{\sigma}^2_{Kernel} = \frac{2}{T}\sum_{j\in J^*}\frac{\triangle K_j}{K_j^2}e^{RT}\hat{Q}(K_j)+\frac{2}{T}ln(\frac{F}{K_0})-\frac{2}{T}[\frac{F}{K_0}-1]. \tag{2.32}$$

Here $J^* = \{j : K_j$ is the j-th strike price, which is either one of the previously chosen one or one of the interpolated or extrapolated strike prices$\}$. We use $\hat{\sigma}^2_{Kernel}$ to estimate the VIX.

Theorem 3.2.1 *The kernel smoothed VIX estimator in equation (2.32) converges to the true theoretical VIX formula in formula (2.27) weakly.*

Proof: We first show that when window width $h \to 0$ and the sample size n is large enough to make sure $nh \to \infty$, the kernel estimator $\hat{Q}$ converges weakly to the true Q[9]. Note, for any fixed x,

$$\begin{aligned} \hat{m}(x) &= \frac{\sum_{i=1}^n K(\frac{x-K_i}{h})\tilde{Q}_i}{\sum_{i=1}^n K(\frac{x-K_i}{h})} \qquad (2.33)\\ &= \frac{\sum_{i=1}^n K(\frac{x-K_i}{h})\left\{m(x)+\Big(m(K_i)-m(x)\Big)+\sigma(K_i)\varepsilon_i\right\}}{\hat{f}(x)} \\ &= m(x)+\frac{\hat{m}_1(x)}{\hat{f}(x)}+\frac{\hat{m}_2(x)}{\hat{f}(x)}, \end{aligned}$$

where

$$\hat{m}_1(x) = \frac{1}{nh}\sum_{i=1}^n K(\frac{x-K_i}{h})\Big(m(K_i)-m(x)\Big),$$

$$\hat{m}_2(x) = \frac{1}{nh}\sum_{i=1}^n K(\frac{x-K_i}{h})\sigma(K_i)\varepsilon_i.$$

[9]Please see Bierens (1987) and Devroye (1978) for related theoretical results. We note that stronger converge results exist for N-W estimator, but there will be a price to pay.

Because $E[\hat{m}_2(x)] = E[E[\hat{m}_2(x)|X]] = 0$,

$$
\begin{aligned}
Var(\hat{m}_2(x)) &= E[Var(\hat{m}_2(x)|K)] + Var(E[\hat{m}_2(x)|K]) \\
&= E[\frac{n}{n^2h^2}K^2(\frac{x-K_i}{h})\sigma^2(K_i)] + 0 \\
&= \frac{1}{nh^2}\int K^2(\frac{x-z}{h})\sigma^2(z)f(z)dz \\
&= \frac{i}{nh^2}\int K^2(v)\sigma^2(x+hv)f(x+hv)dv \\
&= \frac{1}{nh}\int K^2(v)\sigma^2(x)f(x)dv + o(\frac{1}{nh}) \\
&= \frac{\sigma^2(x)f(x)R(K)}{nh} + o(\frac{1}{nh}).
\end{aligned}
$$

Hence, by central limit theorem,

$$
\sqrt{nh}\hat{m}_2 \xrightarrow{\mathbb{D}} N(0, \frac{\sigma^2(x)f(x)R(K)}{nh}). \tag{2.34}
$$

Similarly, work out the expectation and variance of $\hat{m}_1(x)$ and $\hat{f}(x)$ by change of variable and Taylor expansion, apply Law of large number, we get

$$
\hat{f}(x) \xrightarrow{p} f(x), \text{ and} \tag{2.35}
$$

$$
\sqrt{nh}\Big(\hat{m}_1(x) - \frac{1}{2}h^2\frac{m^{"}(x)f(x)+2m^{'}(x)f^{'}(x)}{f(x)}\sigma_K^2\Big) \xrightarrow{p} 0. \tag{2.36}
$$

Hence, by (2.33), (2.36), (2.35), and (2.34),

$$
\sqrt{nh}\Big(\hat{m}(x) - m(x) - \frac{1}{2}h^2\frac{m^{"}(x)f(x)+2m^{'}(x)f^{'}(x)}{f(x)}\sigma_K^2\Big) \xrightarrow{\mathbb{D}} N\Big(0, \frac{\sigma^2(x)R(K)}{nhf(x)}\Big). \tag{2.37}
$$

Based on the weak convergence of $\hat{Q}(n,h)$ to Q, by continuous mapping theorem and sandwich theorem for convergence in probability, the limit can be taken into the sum. And the discretization error goes away when the strike price increment $sup_j(\triangle K_j) \to 0$. The truncation error will be removed when we let $K_L \to 0$ and $K_U \to \infty$. Thus the declaration holds. □

Because the kernel smoothing helps to remove the bias caused by option pricing error, more estimated option prices become available for different strike prices, it's obvious that kernel smooth will help to reduce the corresponding estimation biases introduced by option pricing error and discretization. Our simulation results in the next section justifies the application of our method.

3.3 Simulation Results

In this section, we first present some simulation results case by case to show the bias reduction effect. Then we list some more simulation results based on Black-Scholes Model and Merton Jump Model under a series of parameter specifications. We explore the influence of these parameters on the estimation result.

3.3.1 Comparison of VIX Estimation under Different Methods

To compare the original CBOE method, the spline smoothing method, the kernel smoothing method, and the kernel smoothing with expansion and truncation correction, we illustrate four examples below. Two are without option pricing error and the other two are with pricing error. We provide the scatter plot of estimated option prices that is going to be used for calculating VIX for each example.

In the first two examples, we adopt two specifications. That is time to expire is 30 days, the range of the exercise price is from \$950 to \$1050, the gap between each exercise price is either \$10 or \$5, the underlying index value is assumed 1000. The interest rate is adopted as 0, while the true Black-Schole volatility is set to be 0.2.

From Figure (3.5) and (3.6) respectively, we can see all four methods are under estimating the volatility in these two case. The original CBOE method reports 18.39 and 18.12, the spline smoothing method tells 18.07 and 17.96, the kernel smoothing method gives out 17.98 and 17.84, and the kernel smoothing with expansion and truncation correction improved to be 18.76 and 18.91. The major reason the original CBOE method is underestimating the volatility is because the truncation error. The spline and kernel smoothing method without tail amendment is further underestimating the volatility. The reason is the kernel method we applied here smooths the whole strike vs option price curve, and hence the spike of the option price at the forward price is under estimated. After we extrapolating, extend 25% of the range of the strike price on the each tail to do the prediction, we can see the result is obviously

Figure 3.5 VIX Estimation When No Option Pricing Error Presents; Strike Interval 10.

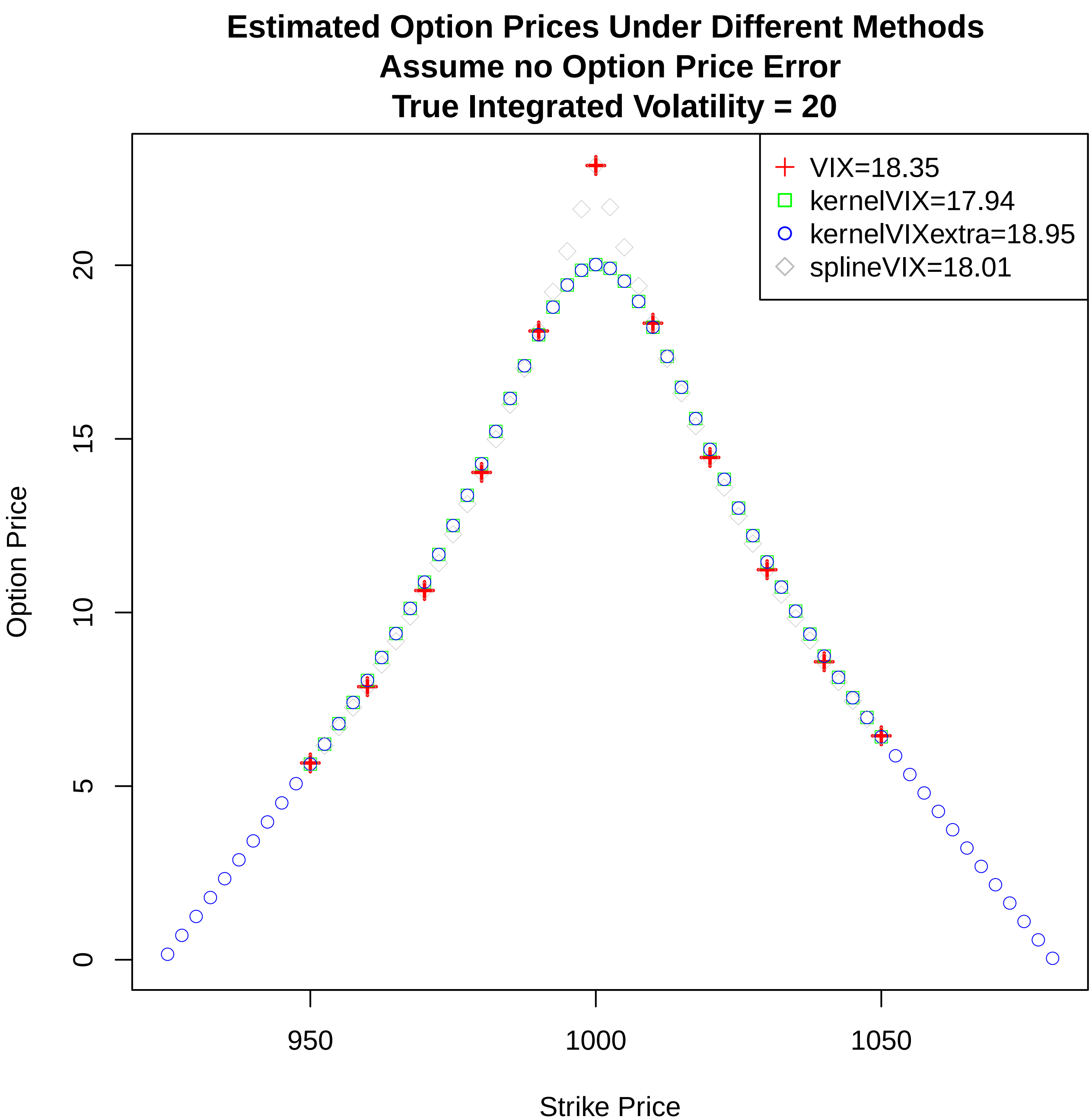
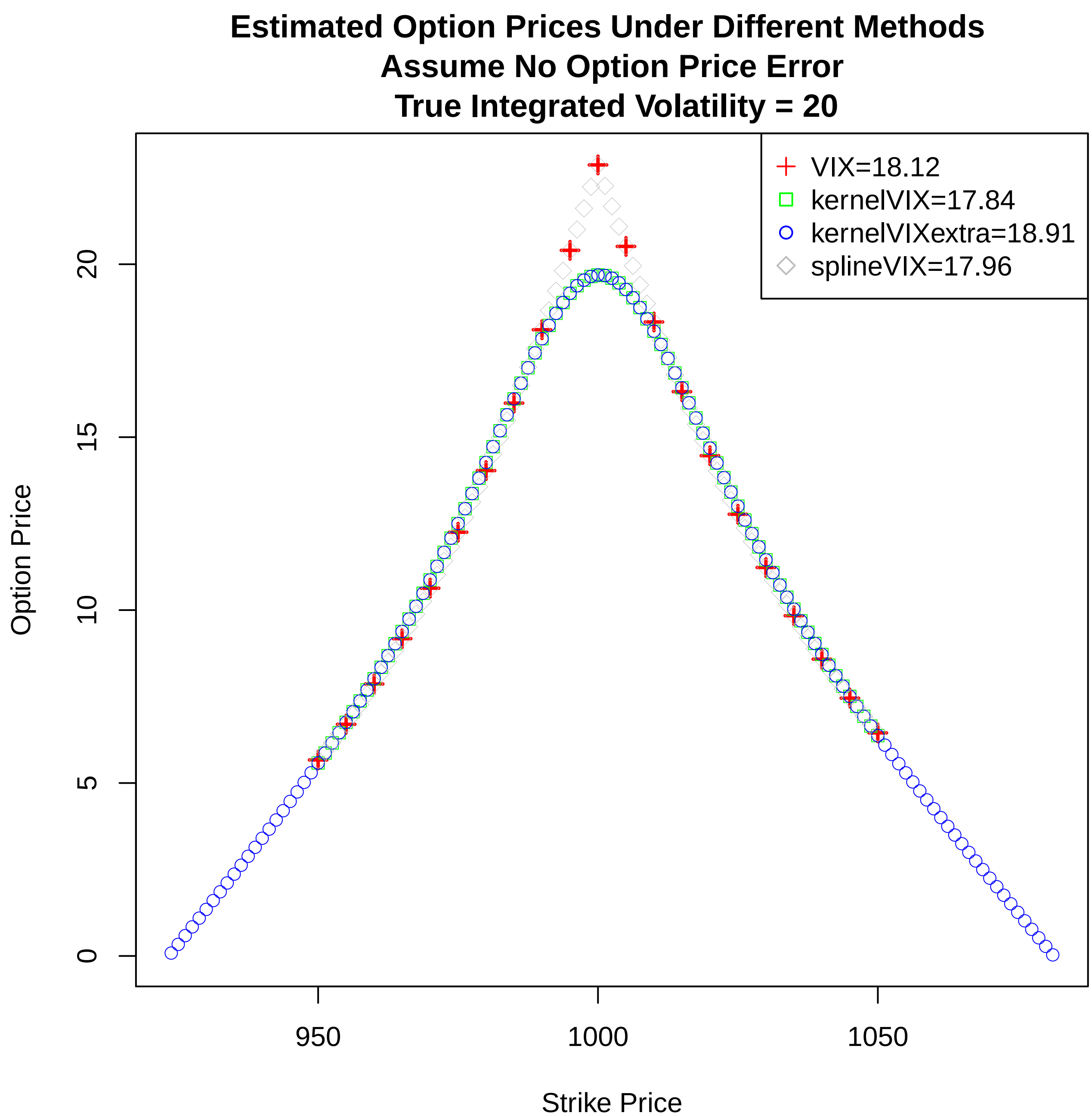


Figure 3.6 VIX Estimation When No Option Pricing Error Presents; Strike Interval 5.



improved.

The option price adopted in the CBOE VIX calculation is the mid-price of the bid and ask. Hence, we have good reason to assume there is pricing error for the options. Because the near-the-money options are with more liquidity, we have reason to believe the pricing error there is smaller than the pricing error of those far out of the money ones. Hence, one error we considered and adopted for the other two examples plots is

$$N * Q * \sqrt{0.01 + 0.3 * ((1 - |(K/S)^{1_{Call} - 1_{Put}} - 1.325|/0.3)_+)^{0.5}}.$$

Here, N denotes random number generated from normal distribution and Q is the option price, K is the strike price, S is the current stock price. When such pricing errors present, we can see from figure 3.7 and 3.8 the VIX estimated under this specification is still underestimated. However, as before, we can see the kernel smoothing method with truncation and expansion correction performs best out the four. This method improves the estimate of VIX from 18.39 to 18.76 when gap is \$10 and from 18.18 to 18.77.

Based on what we observe, the kernel smoothing method with truncation and expansion correction helps reducing the tail truncation error, though with the disadvantage of underestimate the option price at the peak of the curve where the strike price equals the forward price. We also note the original CBOE method though with big disadvantage of truncation error, the discretization which lead to possible over estimation of VIX, also helps to make balance of these errors. Based on observations so far, we may want to advocate our new kernel smoothing method with truncation and expansion correction when sample with very large is not available.

3.3.2 Simulation under Black-Scholes Model and Merton Jump Model

In this subsection, we present some simulation results to continue comparing these different VIX estimation method and try to explore how the estimation result can be influenced by the parameter specifications.

Figure 3.7 VIX Estimation When Option Pricing Errors Presents; Strike Interval 10.

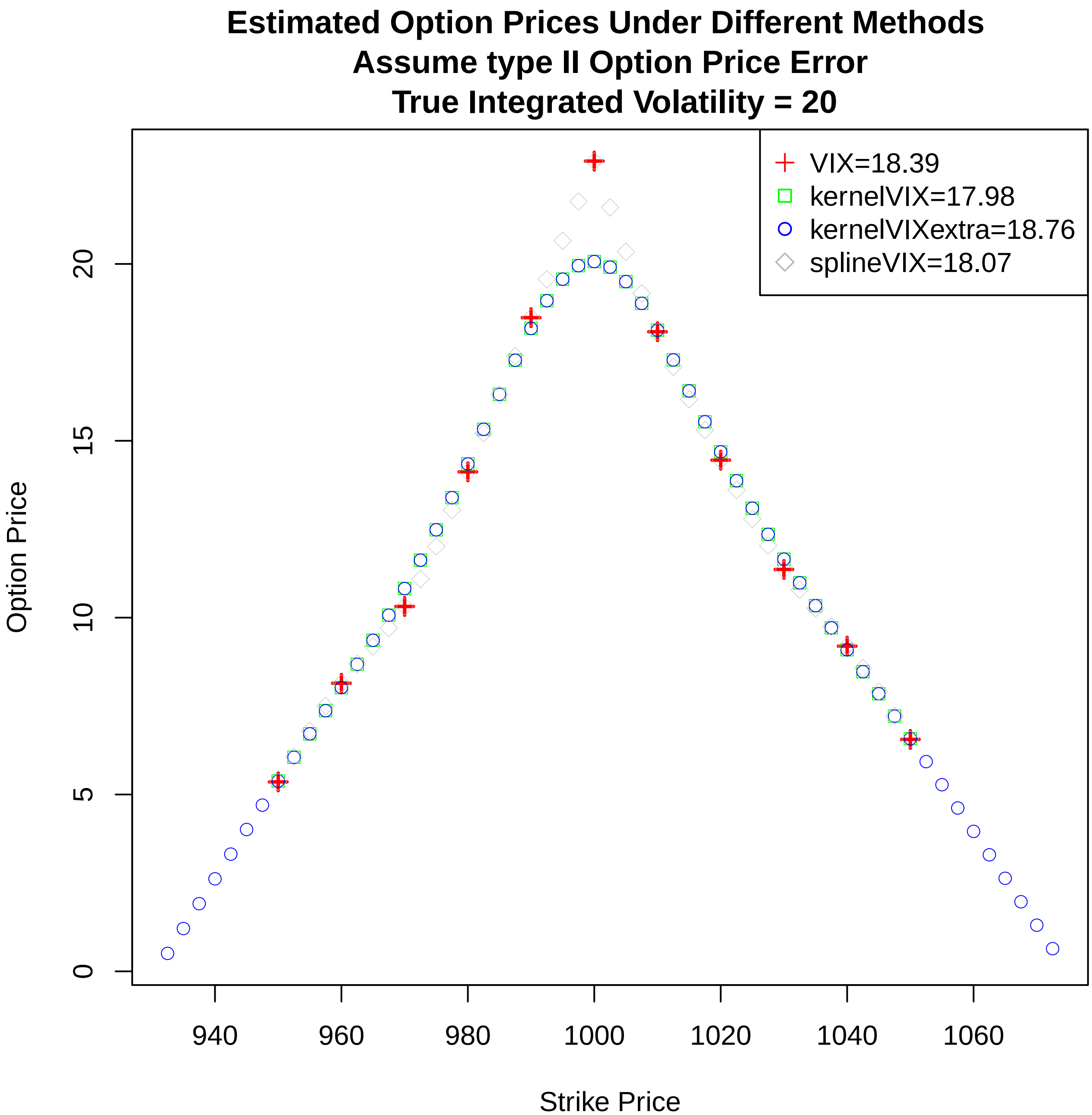
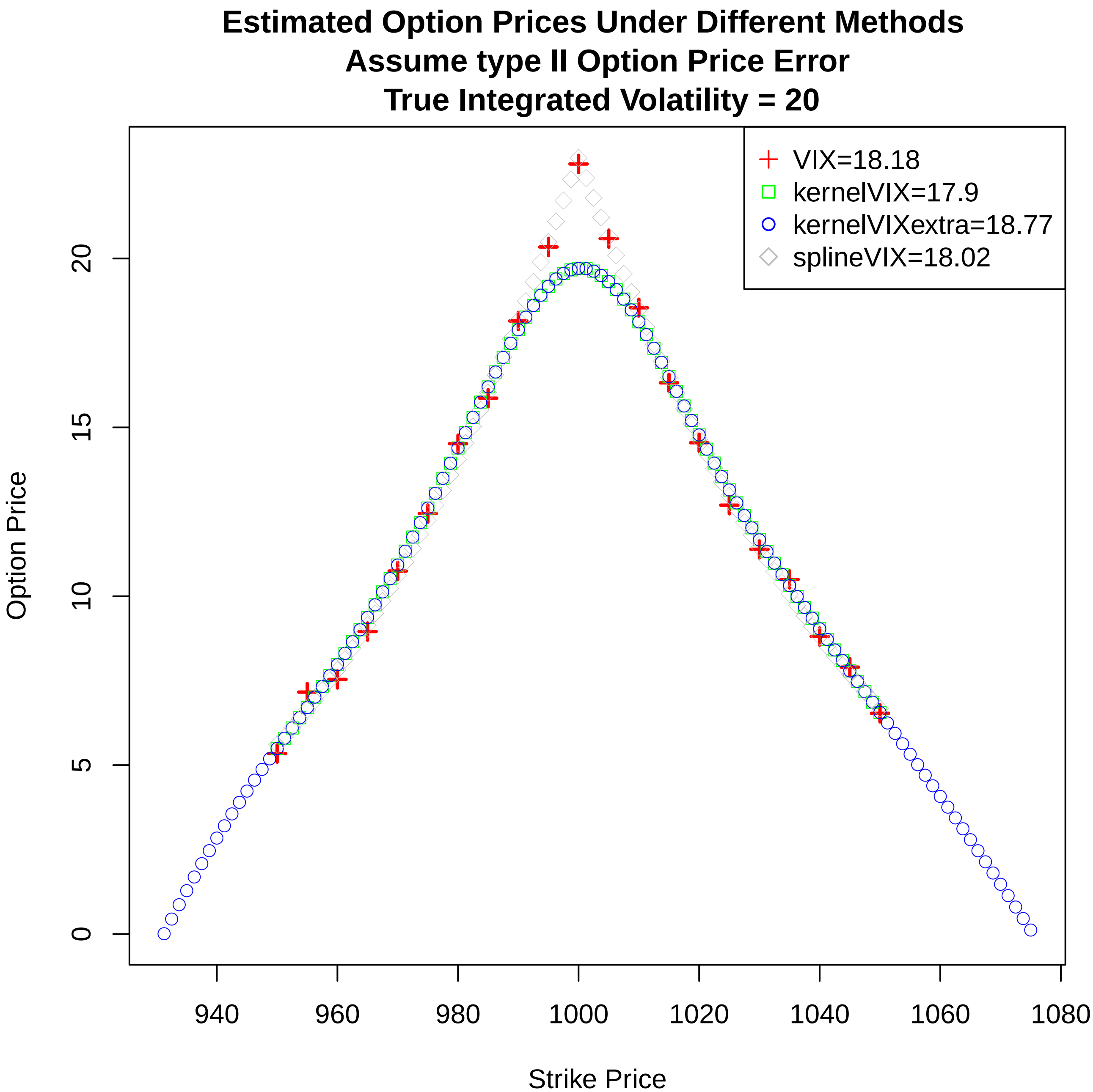


Figure 3.8 VIX Estimation When Option Pricing Errors Presents; Strike Interval 5.



We include Black-Scholes model and Merton jump model. For each model, errors of the following form is added to the theoretical option prices:

$$N(\mu,\sigma^2) * Q * \sqrt{0.01 + 0.3 * ((1 - |(K/S)^{1_{Call}-1_{Put}} - 1.325|/0.3)_+)^{0.5}}.$$

Here, Q denotes the theoretical option prices, $N(\mu,\sigma^2)$ denotes random number generated from normal distribution with mean μ and standard deviance σ_2.

For each model, 36 different parameter specifications are adopted. That is three time to expiration lengths are considered: 15, 30, and 45 days; four combinations of lowest and the highest strike price (KL,KU) are specified: (950,1050), (900,1100), (800, 1200), (700,1300); the strike prices are uniformly distributed between the (KL,KU) with three spaces: 25, 10 and 5.

The interest rate is not playing a very big roll here because we are assuming constant interest rate. Therefore, zero interest rate is assumed. Based on historical volatility estimation for S & P 500, volatility $\sigma = 0.2$ is adopted in both models. The initial underlying price is set to be 1000. For the Merton jump model, we also specify the following parameters based on previous empirical studies to be:

$$\lambda = 0.2, \kappa = 0, \delta = \sqrt{0.15}\sigma.$$

The simulation results are summarized in Table 3.7 and 3.8. TtE denotes days to expire, dK is the space $\triangle K$ between two consecutive strike prices, KL and KU are the minimal and maximal strike price adopted respectively. The corresponding estimation results under our k-ernel smoothed VIX estimator, VIX calculated with more estimated option prices by spline method, and the original VIX formula are denoted as $Kernel$, $Spline$, and VIX. The corresponding estimation variance are indicated by $sdKernel$, $sdSpline$, and $sdVIX$ respectively.

We can see from the simulation result that the kernel smoothed VIX estimation are generally better when error presents. That is partially due to the fact that the option pricing errors are been smooth out. This is also due to the fact that we used more points from the smoothed curve, and hence the discretization errors are reduced. We also expand our smoothed line a little to address the truncation error problem. When time to expire, $\triangle K$ are fixed, the estimated values increases when the range of $[KL, KU]$ increases. From the first four rows of Table 3.7, we can see overestimation occurs when truncation error is reduced with larger strike price range, while large strike intervals are specified. When time to expire and the strike price range are hold constant, the decrease of $\triangle K$ make more option prices available and hence less discretization error. Hence, depends on how severe the truncation error is, this may either help to make the final result more accurate or off the target.

For a typical trading day, the available strike prices are 20 to 40. Hence, as long as we have such moderate number of options, we are not cursed with sparsity and hence the kernel smoothed VIX estimator can provide a better estimation. If for some very extreme cases when sample option numbers are too limited, like for the first case in Table 3.7 where only five option prices are observed, we suggest using the original VIX.

Our method also works for combing several days' options for estimation. For future work, we suggest explore possible explicit fancy relation of options with different expirations, convert them into one same day and then apply kernel smooth. This may help improve the result a lot than simple combination especially when single day strike prices are few for some cases.

Table 3.7 VIX Estimation under Black Scholes Model With Option Pricing Errors

TtE	dK	KL	KU	Kernel	Spline	VIX	sdKernel	sdSpline	sdVIX
15	25	950	1050	20.33	19.85	18.84	0.4757	0.42030	0.74150
15	25	900	1100	20.66	20.59	19.66	0.6619	0.60656	1.08109
15	25	800	1200	20.63	20.66	19.50	1.0759	1.01878	1.95689
15	25	700	1300	20.59	20.69	20.01	1.3971	1.33478	2.29261
15	10	950	1050	19.47	19.15	19.13	0.3220	0.30392	0.32690
15	10	900	1100	20.10	19.98	19.73	0.4425	0.42962	0.66901
15	10	800	1200	20.12	20.12	19.69	0.6456	0.63513	1.04423
15	10	700	1300	20.15	20.23	19.50	0.8641	0.85628	1.64381
15	5	950	1050	19.26	19.00	19.12	0.2276	0.22295	0.23084
15	5	900	1100	20.01	19.87	19.68	0.3044	0.30230	0.50789
15	5	800	1200	20.03	20.04	19.40	0.4652	0.46532	0.76861
15	5	700	1300	20.00	20.07	19.19	0.6359	0.63459	0.82185
30	25	950	1050	19.15	18.41	17.99	0.2524	0.22646	0.30721
30	25	900	1100	20.22	20.03	19.58	0.3381	0.31189	0.43585
30	25	800	1200	20.32	20.32	19.49	0.5446	0.51689	1.16735
30	25	700	1300	20.31	20.39	19.69	0.7048	0.67442	1.34919
30	10	950	1050	18.36	17.95	18.00	0.1708	0.16206	0.16601
30	10	900	1100	19.88	19.67	19.73	0.2237	0.21812	0.25896
30	10	800	1200	20.06	20.03	19.61	0.3236	0.31899	0.69681
30	10	700	1300	20.08	20.14	19.50	0.4317	0.42819	1.10250
30	5	950	1050	18.12	17.84	17.96	0.1209	0.11870	0.11959
30	5	900	1100	19.80	19.61	19.72	0.1538	0.15323	0.18567
30	5	800	1200	20.02	19.98	19.39	0.2329	0.23346	1.17029
30	5	700	1300	20.00	20.06	19.37	0.3175	0.31714	0.46498
45	25	950	1050	18.29	17.45	17.20	0.1762	0.15926	0.20311
45	25	900	1100	19.89	19.61	19.36	0.2292	0.21238	0.28070
45	25	800	1200	20.21	20.17	19.65	0.3648	0.34700	0.74958
45	25	700	1300	20.21	20.27	19.64	0.4717	0.45184	1.02298
45	10	950	1050	17.50	17.07	17.12	0.1194	0.11363	0.11519
45	10	900	1100	19.58	19.34	19.45	0.1514	0.14791	0.15713
45	10	800	1200	20.03	19.96	19.74	0.2160	0.21338	0.45748
45	10	700	1300	20.05	20.10	19.66	0.2892	0.28710	1.14952
45	5	950	1050	17.25	16.97	17.07	0.0847	0.08319	0.08334
45	5	900	1100	19.50	19.29	19.44	0.1041	0.10385	0.10960
45	5	800	1200	20.01	19.92	19.42	0.1554	0.15605	1.42626
45	5	700	1300	20.00	20.04	19.57	0.2116	0.21158	0.54401

Table 3.8 VIX Estimation under Merton Jump Model With Option Pricing Errors

TtE	dK	KL	KU	Kernel	Spline	VIX	sdKernel	sdSpline	sdVIX
15	25	950	1050	20.31	19.83	18.86	0.4797	0.4770	0.6354
15	25	900	1100	20.64	20.57	19.03	0.4813	0.4813	1.0188
15	25	800	1200	20.66	20.69	18.56	0.4898	0.4911	1.4494
15	25	700	1300	20.66	20.69	18.84	0.4737	0.4747	1.3097
15	10	950	1050	19.46	19.14	19.12	0.3672	0.3587	0.3651
15	10	900	1100	20.10	19.98	19.87	0.3622	0.3615	0.3384
15	10	800	1200	20.09	20.10	19.83	0.3507	0.3513	0.3111
15	10	700	1300	20.11	20.14	19.97	0.3594	0.3602	0.3558
15	5	950	1050	19.26	19.00	19.12	0.2631	0.2574	0.2582
15	5	900	1100	20.00	19.87	19.93	0.2449	0.2443	0.2121
15	5	800	1200	20.02	20.03	19.94	0.2618	0.2622	0.2021
15	5	700	1300	20.02	20.06	19.97	0.2670	0.2674	0.2680
30	25	950	1050	19.15	18.41	18.02	0.4931	0.4746	0.4764
30	25	900	1100	20.21	20.02	19.48	0.4969	0.4955	0.5509
30	25	800	1200	20.33	20.33	19.45	0.5142	0.5148	0.6799
30	25	700	1300	20.34	20.42	19.39	0.4866	0.4891	0.7433
30	10	950	1050	18.36	17.95	18.00	0.3692	0.3560	0.3648
30	10	900	1100	19.88	19.67	19.73	0.3591	0.3563	0.3362
30	10	800	1200	20.04	20.00	19.91	0.3444	0.3452	0.2766
30	10	700	1300	20.05	20.12	19.91	0.3508	0.3523	0.2755
30	5	950	1050	18.12	17.84	17.95	0.2619	0.2551	0.2595
30	5	900	1100	19.80	19.60	19.73	0.2428	0.2401	0.2278
30	5	800	1200	20.01	19.97	19.95	0.2569	0.2573	0.2063
30	5	700	1300	20.01	20.07	19.94	0.2630	0.2640	0.1920
45	25	950	1050	18.30	17.46	17.23	0.4929	0.4646	0.4494
45	25	900	1100	19.89	19.61	19.35	0.5010	0.4971	0.5017
45	25	800	1200	20.22	20.17	19.69	0.5211	0.5213	0.5356
45	25	700	1300	20.24	20.29	19.66	0.4888	0.4910	0.5279
45	10	950	1050	17.50	17.07	17.12	0.3636	0.3484	0.3570
45	10	900	1100	19.58	19.34	19.44	0.3559	0.3514	0.3413
45	10	800	1200	20.01	19.94	19.92	0.3358	0.3365	0.2776
45	10	700	1300	20.03	20.08	19.93	0.3423	0.3438	0.2677
45	5	950	1050	17.25	16.97	17.06	0.2565	0.2497	0.2539
45	5	900	1100	19.50	19.29	19.43	0.2414	0.2375	0.2344
45	5	800	1200	20.00	19.92	19.94	0.2514	0.2516	0.2100
45	5	700	1300	20.00	20.04	19.94	0.2580	0.2590	0.1956

CHAPTER 4. GENERAL CONCLUSION

In this thesis, we have developed some new methods to do certain statistical estimations when the underlying processes are specified by stochastic diffusion equations.

In Chapter 2, we work under general stochastic diffusion processes and general estimating equations to estimate model parameters. We use tensor method to expand the large class of estimators defined by estimating equations, which consist maximum likelihood estimators, method of moments estimators, approximate likelihood estimators, and so on. With the aid of the central limit theory for strong mixing processes, we are able to conducted calculation of the conditional expectations using infinitesimal generators under some regularity conditions when implicit transitional density of the original process is unknown. We have theoretically quantified the high order bias and variance of the drift and diffusion parameters. For mean reversion processes like two dimensional O-U process, we have derived that the estimation bias for the mean reversion rates are of order $O(\frac{1}{T})$, while the order of the estimation bias for the volatilities are $O(\frac{1}{n})$ and $o(\frac{1}{n})$ for the long term means. This helps us to understand the reason we see much larger biases in the estimated mean reversion rates than other parameters. Simulations are conducted to compare our theoretical results and the numerical results. Parametric bootstrap method is applied in some simulation cases and verifies our assumption that bootstrap can help with bias reduction dramatically when sample sizes are relatively small.

Our Chapter 3 reviewed the volatility index estimator, VIX, suggested by Chicago Board Option Exchange and Goldman Sachs. Their estimator is derived based on the concept of fair value of future variance and is essentially a weighted sum of the near-the-money option prices for both call and put options. In order to reduce the discretization error, truncation error, and the approximation error in their estimator, as well as to reduce the estimation bias introduced by option pricing errors, we propose a new method which combines the CBOE method and the kernel smoothing method. Our new estimator converges weakly to the true integrated volatility. Some simulations are performed assuming Black-Scholes model and Merton model with jumps to compare different estimators. Our estimator improves the estimation results under our targeted cases.

BIBLIOGRAPHY

References

Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. *Econometrica.* **70**, 223-262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics.* **36**, 906-937.

Aït-Sahalia, Y., Cacho-Diaz, J. and Hurd, T. (2009). Portfolio choice with a jumps: a closed form solution. *Annals of Applied Probability.* **19**, 556-584.

Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics.* **83**, 413-452.

Applebaum, D. (2005). *Lévy processes and stochastic calculus.* University press, Cambridge.

Bertoin, J. (1996). *Lévy processes.* University press, Cambridge.

Billingsley, P. (1968.) *Convergence of probability measures.* John Wiley and Sons.

Bibby, M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli.* **1**, 17-39.

Bierens, H.(1987). Kernel estimators of regression functions. *Truman F. Bewley (ed.), Advances in Econometrics: Fifth World Congress, Vol. I, Cambridge University Press.* 99-144.

Bessembinder, H., Coughenour, J., Seguin, P. and Smoller, M. (1995). Mean reversion in equilibrium asset prices: evidence from the futures term structure. *Journal of Finance.* **1**, 361-375.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy.* **81**, 637-659.

Bollen, N. and Whaley, R. (2004.) Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance.* **59**, 711-754.

Bosq, D. (1998). *Nonparametric statistics for stochastis processes.* Springer-Verlag, New York.

Breeden, D. and Litzenberger, R. (1978). Prices of state-contigent claims implicit in option prices. *Journal of Business.* **51**, 621-651.

Britten-Johns, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance.* **2**, 839-866.

Broadie, M. and Jain, A. (2008). The effect of jumps and discrete sampling on volatility and variance swaps. *International Journal of Theoretical and Applied Finance.* **11**, 761-797.

CBOE (2009). *The CBOE Volatility Index - VIX; CBOE White Paper.* From http://www.cboe.com/micro/vix/vixwhite.pdf.

Chen, S. and Cui, H. (2007). On the second-order properties of empirical likelihood with moment restrictions. *Journal of Econometrics.* **141**, 492-516.

Cox, J., Ross, S. and Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics.* **7**, 229-263.

Cox, J., Ingersoll, J. and Ross, S. (1985). A theorem of the term structure of interest rates. *Econometrica.* **53**, 385-407.

Demeterfi, K., Derman, E., Kamal, M. and Zou, J. (1999). A guide to volatility and variance swaps. *Journal of Derivatives.* **6**, 9-32.

Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *The Canadian Journal of Statistics.* **6**, 179-191.

Dacunha-Castelle, D. and Duflo, M. (1986). *Probability and Statistics: Vol. 2.* Springer.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli.* **6**, 1051-1079.

Hansen, L. and Scheinkman, J. (1995). Back to the future: generating moment implications for continuous time Markov processes. *Econometrica.* **63**, 767-804.

Hardin, J. and Hilbe, J. (2003). *Generalized estimating equations.* Chapman and Hall.

Härdle, W. (1989). *Applied nonparametric regression.* Cambridge University Press.

Harvey, R. and Whaley, R. (1992). Dividends and S & P 100 index option valuation. *Journal of Futures Markets.* **43**, 123-138.

Hull, J. (2005). *Options, futures and other derivatives.* Prentice Hall.

Jacod J. and Shiryaev, A. (1987). *Limit theorems for stochastic processes.* Springer-Verlag, Berlin.

Jiang, G. and Tian, Y. (2005). The model-free implied volatility and its information content. *Review of Financial Studies.* **18**, 1305-1342.

Jiang, G. and Tian, Y. (2007). Extracting model-free volatility from option prices: an examination of the VIX index. *The Journal of Derivative.* **14**, 35-60.

Joshi, M. (2008) *The concepts and practice of mathematical finance.* Cambridge University Press.

Karatzas, I. and Shreve, S. (1988). *Brownian motion and stochastic calculus.* Springer.

Krumscheid, S., Pavliotis, G. and Kalliadasis, S. (2013). Semiparametric drift and diffusion estimation for multiscale diffusions. *Multiscale Modeling and Simulation.* **11**, 442-473.

Lamberton, D. and Lapeyre, B. (1996), *Introduction to Stochastic Calculus Applied to Finance.* Chapman and Hall.

Merton, R. (1973). The theory of rational option pricing. *Bell Journal of Economics and Management Science.* **4**, 141-183.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics.* **3**, 125-144.

McCullagh, P. (1987) *Tensor methods in statistics.* New York : Chapman and Hall.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications.* **9**, 141-142.

Nkurunziza S. (2013) Preliminary test and estimation in some multifactor diffusion processes. *Sankhyā: The Indian Journal of Statistics, Series A.* **75**, 211-230.

Nowman, K. (1997). Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *Journal of Finance.* **52**, 1695-1706.

Philips, P. (1972). The structural estimation of stochastic differential equation systems. *Econometrica.* **40**, 1021-1041.

Protter, P. (2004). *Stochastic integration and differential equations.* Springer, Berlin.

Philips, P. and Yu, J. (2005). Jackknifing bond option price. *Review of Financial Studies.* **18**, 707-742.

Philips, P. and Yu, J. (2009). Simulation-based estimation of contingent-claims prices. *Reviews of Financial Studies.* **22**, 3669-3705.

Øksendal, B. (1998). *Stochastic differential equations.* Springer-Verlag, Berlin Heidelberg.

Rao, B. (1976). *Normal approximation and asymptotic expansions.* John Wiley and Sons.

Sargan, J. (1976). Econometric estimators and the edgeworth approximation. *Econometrica.* **44**, 421-448.

Serfling, R. (1980). *Approximation theorems of mathematical statistics.* John Wiley and Sons.

Simmonds, J. (1994). *A brief tensor analysis.* Springer-Verlag, New York.

Song, Z. and Xiu, D. (2012). A tale of two option markets: state-price densities implied form S & P 500 and VIX option prices. *Chicago Booth Research Paper No. 12-10.*

Stoll, H. (1969). The relationship between put and call option prices. *The Journal of Finance.* **24**, 801-824.

Tang, C. and Chen, S. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics.* **149**, 65-81.

Varughese, M. (2013). Parameter estimation for multivariate diffusion systems. *Computational Statistics & Data Analysis.* **57**, 417-428.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics.* **5**, 88-177.

Wang, X., Philips P. and Yu, J. (2011). Bias in estimating multivariate and univariate diffusions. *Journal of Econometrics.* **161**, 228-245.

Watson, G. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A.* **26**, 359-372.

Whaley, R. (1993). Derivatives on market volatility: hedging tools long overdue. *The Journal of Derivatives.* **1**, 71-84.

Whaley, R. (2009). Understanding the VIX. *The Journal of Portfolio Management.* **35**, 98-105.

Yu, J. (2007). Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese Yuan. *Journal of Econometrics.* **141**, 1245-1280.